ELDAR SHAFIR, EDITOR

The Behavioral Foundations of Public Policy

The Behavioral Foundations of Public Policy

The Behavioral Foundations of Public Policy

EDITED BY ELDAR SHAFIR

Contents

Foreword

DANIEL KAHNEMAN

There are no established churches in the Woodrow Wilson School of Public and International Affairs at Princeton, but there have always been established disciplines. Originally there were two: economics and politics (elsewhere known as political science). In 1999, psychology was formally introduced as the third discipline, and granted the intimidating responsibility for a semester-long compulsory class to all students working toward the degree of master of public affairs. We[1] had to find answers to some difficult questions: What does psychology have to offer to students who prepare for a career of public service? What gaps existed in our students' training that we should fill? What biases in their training should we aim to correct?

The question about biases was the easiest to answer. We observed that the students in the master's program offered by the School were exposed to a steady diet of economics courses that invoked the standard assumption of agents who are invariably rational, driven by self-interest, and motivated by tangible incentives. In the eyes of a psychologist, these propositions are not viable even as a crude approximation. The tension between psychology and the assumptions of economic theory provided a natural focus for the course we designed. Accordingly, our course emphasized errors of judgment, oddities of choice, the power of framing effects, and the intense and universal concern of people with their social group and their standing within it. We wanted our students to know that the assumptions of the rational agent model, although adequate for predicting the outcomes in many markets, are not at all adequate for predicting how individuals will actually behave in most situations. The policy-relevant situations we explored extended beyond purely economic circumstances, to issues ranging from voting and negotiations, to health behaviors, labor relations, education, and the law.

So why focus on economics in a course on psychology and policy, or in the foreword of a book about that subject? Like it or not, it is a fact of life that economics is the only social science that is generally recognized as relevant and useful by policy makers. Given their monopoly, economists have become gatekeepers, and their analyses and conclusions have

enormous weight even in domains in which they do not seem to have any particular comparative advantage, such as health care and education. An obvious asymmetry in the distribution of competence contributes to the elevated status of economics: there are important policy questions that only economists are qualified to answer, but hardly any data of other social sciences that they cannot evaluate. In particular, economists have more statistical tools at their disposal than most other social scientists do. Even more important, they are native to the universal language of policy, which is money. Finally, their reputation for hard-headed objectivity gives them a significant credibility advantage over more tender-hearted practitioners of the social sciences, whom I have heard casually dismissed as "social workers."

We considered our Princeton policy students as future policy makers, who would be exposed to economic approaches to all fields of social policy. Our intent was to sensitize them to the potential pitfalls of basing policy on the standard assumptions of the rational agent model. We also mentioned to them that a growing minority of economists—behavioral economists—were engaged in attempts to develop an economic science that is based on more realistic psychological assumptions. Behavioral economics was at the time clearly defined as a distinctive approach to economics, with no particular applications to policy.

The landscape changed radically during the first decade of the new century. Behavioral economists began to address the world at large, and the boundary between behavioral economics and applied social psychology blurred, creating a new set of problems and opportunities for psychologists interested in policy. In 2001 Richard Thaler and Shlomo Benartzi reported on the success of their now famous Save More Tomorrow method for increasing workers' willingness to save from their salary. They identified three psychological obstacles to saving: loss aversion, hyperbolic discounting, and status quo bias. Save More Tomorrow was an offer to workers that bypassed these obstacles, leading them to save more. The same year, Bridget Madrian and Dennis Shea published a paper showing that an even simpler procedure—merely

changing the default—can help increase enrollments in savings plans. Now, a decade later, automatic enrollment and automatic escalation (a generic form of Save More Tomorrow) are affecting the lives and savings decisions of millions of people around the world.

A social psychologist will recognize both these strategies as brilliant reinventions of the classic Lewinian proposal for inducing behavioral change, which favors reducing the "restraining forces" over increasing the "driving forces." To follow the Lewinian approach one begins by asking "why don't people already do what I wish they would do?" This question evokes a list of restraining forces, which the agent of change then works to reduce or eliminate. The idea is transparently correct when you are exposed to it, but it is also deeply counterintuitive. The standard tools that most of us use to change others' behavior are arguments, promises, and threats. It is much less natural to look for ways of making it easier for the other person to do the right thing. Thaler and Benartzi developed a procedure that made it easy for the worker to commit to a higher saving rate in the future, which would start automatically at an auspicious time (upon receiving a salary raise). Ending the commitment, in contrast, would require a deliberate decision and a modest effort.

In subsequent articles and in their international best seller, *Nudge*, Thaler and Cass Sunstein described an approach to policy that they called "libertarian paternalism." The central idea is that it is legitimate for institutions of society to consider the best interests of individuals in structuring the choices that these individuals make—for example, about retirement saving. The goal is to make it easy and natural for casual decision makers to make sensible choices, while ensuring their complete freedom to choose as they will. This was read by all as a manifesto of the approach of behavioral economics to policy. It is founded on the ideas that the rational agent model is unrealistic, that many decisions are made with little thought, and that it is appropriate to create a "choice architecture" that reduces the incidence of foolish decisions without reducing freedom.

We have known for a long time that the role of economics in formulating policy has significant consequences. During the heyday of the rational agent model, policies were sometimes formulated that assumed rationality as a psychological fact. For example, the assumption that criminals are rational agents implies that they can be deterred by the expected disutility of being caught and punished. The probability of being caught and the severity of punishment have equivalent weights in this model, but not in reality: empirical research suggests that increasing the probability of punishment is far more effective in deterring

crime than a corresponding increase of severity. In other situations, the rational agent model implies that agents need no protection against their own bad choices: choices freely made by rational agents deserve complete respect. To the surprise of most noneconomists, complete respect is often extended to awful choices, such as those that lead to addiction to noxious substances, or to lives of destitution after retirement. Because psychologists are not trained to assume that humans are rational, they are likely to find this position unattractive and even bizarre—but they recognize the risk that paternalism poses to the ideal of liberty. *Nudge* showed a way out of this dilemma: simple procedures that tend to bias people toward sensible and socially desirable choices without in any way abridging their freedom.

Nudge relied on psychology to highlight another objective that would be pointless if humans were fully rational in the role of consumers. Everyone recognizes that consumers need protection against predatory behavior, and there are many laws that are designed to provide such protection. However, the authors of *Nudge* documented many ways in which firms may take advantage of the psychological limitations of lazy and boundedly rational consumers. The book, along with work by several other researchers, showed how simple regulations can constrain predatory (though not illegal) behaviors, such as formulating truthful contracts in impenetrable language and printing them in painfully small print.

The publication of *Nudge* was immediately recognized as an important event. Sunstein became Director of the Office of Information and Regulatory Affairs (aka the "regulation czar") under President Obama, and Thaler became an advisor to a Behavioral Insight Team (colloquially known as the "Nudge Unit") established by the coalition government led by David Cameron in the UK. Other nudge units are popping up elsewhere around the world with the goal of establishing policies to help people make decisions that serve their best interest, and to protect them from exploitation in the market. The success of this enterprise can be counted as one of the major achievements of applied behavioral science in general, and of applied social and cognitive psychology in particular.

Unfortunately, because the two authors of *Nudge* were an economist and a jurist, respectively the intellectual leaders of behavioral economics and of behavioral law and economics, not only the ideas they produced themselves but also many of the contributions of cognitive and social psychology on which they had relied were labeled "behavioral economics" in the press.[2] And so it came to pass that many applications of social and cognitive psychology came to be called behavioral economics, and many psychologists

discovered that the name of their trade had changed even if its content had not. Quite a few of the authors of chapters in this book would be incorrectly described in the press as behavioral economists, because what they do is develop some of the fundamental theories and document some of the central findings on which *Nudge*, and related writings, have relied. This is not the outcome that most researchers, including the authors of *Nudge*, encouraged or viewed as desirable. Richard Thaler has always insisted on a narrow definition of behavioral economics as a distinctive approach of economics, and he would prefer to see "nudges" described as applications of behavioral science.

Labels matter, and the mislabeling of applied behavioral science as behavioral economics has consequences. Some are positive; behavioral economics has retained the cachet of economics, so psychologists who are considered behavioral economists gain some credibility in the policy and business worlds. But the cost is that the important contributions of psychology to public policy are not being recognized as such, and there is the very real worry that young psychologists will be put off from doing policy-related work because they do not consider themselves economists, even with the modifier "behavioral" as a prefix. It is regrettable that the discipline of psychology gets no credit for the most consequential applications of psychological wisdom, and that students of psychology, who ought to take greater pride in their profession, are left to wonder about the contributions of their discipline to society.

In fact, there is a lot to be done. Nudges are an effective way to use psychological insight in the design of policies that might generate greater welfare. But some policy issues will need a greater rethinking: a questioning of the fundamental assumptions, rather than nuanced design. When it comes to the memories of eyewitnesses, or to employers' ability to avoid discrimination, or to the budgeting challenges of the poor, behavioral research presents the serious possibility that we may want to rethink some fundamental concepts and question the basic assumptions of current policies—in other words, do more than merely nudge.

I hope this book helps steer us in the right direction in giving behavioral scientists a greater role in policy making around the world. The chapters of this book, written predominantly by psychologists, illustrate how much psychology has to offer to policy. An important conclusion that readers should draw from it is that modern psychology has agreed on some important aspects of both human nature and the human condition. Recent years have seen a convergence of views on the roles of cognitive and emotional factors as determinants of behavior—and therefore as targets for policy interventions that are proposed to modify people's circumstances or their actions. There is also a growing recognition of the role of social and cultural drivers of behavior, though many social scientists will still complain that psychology is insufficiently attuned to issues of culture and identity. The recognition of the huge power of situation, context, priming, and construal is common ground. We are all Lewinians now, and in the context of policy behavioral economists are Lewinian as well.

The relationship between psychology and economics in the domain of policy was a central issue when psychology became one of the core disciplines in the Woodrow Wilson School at Princeton. In a very different way it is still a dilemma, not because the disciplines are so alien, but rather because they are so close. The overlap of interests and methods is much greater than it was fifteen years ago. Indeed there are several domains in which members of the different tribes deal with similar problems in similar ways. The study of happiness is one of these domains, the study of inequality and poverty may be another. And there will be more. We need a common label for our shared activities. "Behavioral economics" is not a good label, simply because psychologists are not economists and are not trained to think about markets. "Social psychology" would cause similar difficulties to the economists, lawyers, and physicians who engage in Lewinian practice. A descriptively correct label is "applied behavioral science." I would be proud to be called an applied behavioral scientist, and I believe most of the authors of this book would also be happy to be counted as members of this club. This book is a fine illustration of the potential contribution of applied behavioral science to policy.

Notes

1. "We" refers to myself, Eldar Shafir, and Rob McCoun, who came to help us from the Goldman School of Public Policy at Berkeley.

2. Not only in the popular press. I am on record as describing *Nudge* as "the major accomplishment of behavioral economics." I was quite slow to recognize the problem that I address in this foreword.

Contributors

Adam L. Alter
  Assistant Professor of Marketing, Leonard N. Stern School of Business, New York University
  Adam Alter's research focuses on decision making and social psychology, with a particular interest in the effects of subtle environmental cues on human cognition and behavior.

John P. Balz
  Senior Planner, Draftfcb
  John Balz develops marketing strategies for major global firms. He was introduced to behavioral ideas in graduate school, where his research focused on behavioral decision making in campaigns and elections, congressional lobbying, and ideological attitude formation.

Mahzarin R. Banaji
  Richard Clarke Cabot Professor of Social Ethics, Department of Psychology, Harvard University
  Mahzarin Banaji is interested in the unconscious assessment of self and others, its roots in feelings and knowledge about social group membership, and the perceived distinction between "us" and "them" that is often the result.

Michael S. Barr
  Professor of Law, School of Law, The University of Michigan
  Michael Barr conducts large-scale empirical research on financial services and low- and moderate-income households and writes about a wide range of issues in financial regulation and the law.

Shlomo Benartzi
  Professor, Anderson Graduate School of Management, UCLA
  Shlomo Benartzi's interests lie in behavioral finance with a special interest in personal finance and participant behavior in defined contribution plans.

Paul Brest
  President, William and Flora Hewlett Foundation; Professor Emeritus and Former Dean, Stanford Law School
  Paul Brest has most recently focused on issues of problem solving, judgment, and decision making in the context of the law.

Geoffrey L. Cohen
  James G. March Professor of Organizational Studies in Education and Business; Professor, Department of Psychology, Stanford University
  Geoffrey Cohen's research examines processes related to identity maintenance and their implications for social problems.

William J. Congdon
  Research Director, Economic Studies Program, The Brookings Institution
  William Congdon studies how best to apply behavioral economics to public policy.

John M. Darley
  Warren Professor of Psychology, Department of Psychology, Princeton University
  John Darley's research focuses on moral decision making, decisions to punish another for a transgression, and how people attempt to manage others through incentive systems.

John F. Dovidio
  Professor, Department of Psychology, Yale University
  John Dovidio investigates issues of social power and social relations, with a focus on conscious and unconscious influences on how people think about, feel, and behave toward others based on group membership.

Sara L. Eggers
  Independent Researcher
  Sara Eggers' work focuses on risk communication, stakeholder engagement, and the development and use of decision-analytic tools to support policy making in the environmental and public health arenas.

Phoebe Ellsworth
  Frank Murphy Distinguished University Professor of Law and Psychology, The University of Michigan

Phoebe Ellsworth's research interests lie at the crossroads of psychology and law, particularly person perception, emotion and public opinion, the death penalty, and jury behavior.

Baruch Fischhoff
Howard Heinz University Professor, Department of Social and Decision Sciences, Department of Engineering and Public Policy, Carnegie Mellon University
Baruch Fischhoff's research looks at issues around risk analysis, communication and management, adolescent decision making, informed consent, and environmental protection.

Susan T. Fiske
Eugene Higgins Professor of Psychology, Department of Psychology, Princeton University
Susan Fiske's research addresses how stereotyping, prejudice, and discrimination are encouraged or discouraged by social relationships, such as cooperation, competition, and power.

Craig R. Fox
Professor of Policy, Andersen School of Management; Professor of Psychology, Department of Psychology, UCLA
Craig Fox studies behavioral decision theory, with a focus on how people make judgments and decisions under conditions of risk, uncertainty, and ambiguity.

Julio Garcia
Research Associate, Department of Psychology and Neuroscience, University of Colorado, Boulder
Julio Garcia works with schools to examine processes related to identity maintenance with the aim to reduce racial and gender gaps and improve academic performance.

Alan S. Gerber
Charles C. and Dorathea S. Dilley Professor of Political Science, Yale University
Alan Gerber studies the application of experimental methods to the study of campaign communications. He has designed and performed experimental evaluations of partisan and nonpartisan campaigns and fundraisers.

Daniel G. Goldstein
Principal Research Scientist, Yahoo Research
Daniel Goldstein studies heuristics and bounded rationality as they relate to matters of business and policy, with an emphasis on perceptions of risk and uncertainty, particularly in financial markets.

Ryan Goodman
Anne and Joel Ehrenkranz Professor of Law, New York University School of Law
Ryan Goodman's interests lie in the areas of international human rights law and international relations, focusing on the evaluation of human rights treaties and international law more generally.

Sam Gross
Thomas and Label Long Professor of Law, School of Law, The University of Michigan
Sam Gross' work focuses on the death penalty, false convictions, racial profiling, eyewitness identification, the use of expert witnesses, and the relationship between pretrial bargaining and trial verdicts.

Curtis D. Hardin
Professor, Department of Psychology, CUNY, Brooklyn College
Curtis Hardin's research focuses on the interpersonal foundations of cognition, including the self-concept, social identification, prejudice, and ideology.

Derek Jinks
The Marrs McLean Professor in Law, School of Law, The University of Texas at Austin
Derek Jinks' research and teaching interests include public international law, international humanitarian law, human rights law, and criminal law.

Christine Jolls
Gordon Bradford Tweedy Professor of Law and Organization, Yale Law School
Christine Jolls' research and teaching concentrate in the areas of employment law, privacy law, behavioral law and economics, and government administration.

Leslie John
Assistant Professor of Business Administration, Harvard Business School
Leslie John uses both laboratory and field studies to investigate behavioral questions at the intersection of marketing and public policy concerning consumer privacy and well-being.

Eric J. Johnson
Norman Eig Professor of Business Marketing, Columbia Business School
Eric Johnson's research interests are in consumer and managerial decision making and electronic commerce.

Linda H. Krieger
Professor of Law, William S. Richardson School of Law, University of Hawai'i at Mānoa

Linda Krieger's research focuses on disability discrimination, affirmative action, law and social cognition, judgment in legal decision making, and theories of law and social change.

Howard Kunreuther
James G. Dinan Professor of Decision Sciences and Business and Public Policy, The Wharton School, University of Pennsylvania
Howard Kunreuther's interests involve the ways that society can better manage low-probability/high-consequence events related to technological and natural hazards.

Judith Lichtenberg
Professor of Philosophy, Georgetown University
Judith Lichtenberg focuses on ethics and political philosophy, with special interests in justice and charity, race and ethnicity, and moral psychology.

George Loewenstein
Herbert A. Simon Professor of Economics and Psychology, Department of Social and Decision Sciences, Carnegie Mellon University
George Loewenstein's work brings psychological considerations to bear on models and problems that are central to economics, with a special interest in intertemporal choice.

Elizabeth F. Loftus
Distinguished Professor of Social Ecology, and Professor of Law and Cognitive Science, University of California, Irvine
Elizabeth Loftus studies human memory and examines how facts, suggestions, and other postevent information can modify people's memories.

Robert Meyer
Gayfryd Steinberg Professor of Marketing, The Wharton School, University of Pennsylvania
Robert Meyer's research focuses on consumer decision analysis, sales response modeling, and decision making under uncertainty.

Erwann Michel-Kerjan
Adjunct Associate Professor, Operations and Information Management Department, and Managing Director, Risk Management and Decision Processes Center, The Wharton School, University of Pennsylvania
Erwann Michel-Kerjan's work focuses on strategies for managing the risks, the financial impact, and the public policy challenges associated with catastrophic events.

Dale T. Miller
The Class of 1968 / Ed Zschau Professor of Organizational Behavior, Stanford Graduate School of Business
Dale Miller's research interests include the impact of social norms on behavior and the role that justice considerations play in individual and organizational decisions.

Sendhil Mullainathan
Professor, Department of Economics, Harvard University; Scientific Director and Founder, ideas42
Sendhil Mullainathan conducts research on development economics, behavioral economics, and corporate finance, with a focus on the application of the behavioral perspective to policy.

Ehud Peleg
Head of Enterprise Risk Management, Bank Leumi
Ehud Peleg's interests are in the applications of behavioral science to investment and resource allocation decisions at the individual executive, committee, and board levels.

Deborah A. Prentice
Alexander Stewart 1886 Professor of Psychology, Department of Psychology, Princeton University
Deborah Prentice studies how norms guide and constrain behavior, how people respond when they feel out of step with prevailing norms, and how they react to those who violate social norms.

Emily Pronin
Associate Professor, Department of Psychology, Princeton University
Emily Pronin studies the asymmetries in how we perceive ourselves versus how we perceive others and the ways in which the underlying processes can lead to misunderstanding and conflict.

Donald A. Redelmeier
Canada Research Chair in Medical Decision Sciences; Professor of Medicine, University of Toronto; Director of Clinical Epidemiology, Sunnybrook Health Sciences Centre.
Donald Redelmeier studies decision sciences in medical contexts, particularly the role of judgmental error and the opportunity for improvement in policy and treatment.

Jennifer A. Richeson
Professor, Department of Psychology, Northwestern University
Jennifer Richeson studies the ways in which social group memberships such as race, socioeconomic

status, and gender impact prejudice, stereotyping, and intergroup relations.

Todd Rogers

Assistant Professor of Public Policy, Harvard Kennedy School

Todd Rogers uses the tools and insights of behavioral science to understand and study how to influence socially consequential problems.

Lee Ross

Professor, Department of Psychology, Stanford University

Lee Ross's research focuses on the biases that lead people to misinterpret each other's behavior, thereby creating systematic barriers to dispute resolution and the implementation of peace agreements.

Kathleen Schmidt

Graduate Student, Department of Psychology, The University of Virginia

Kathleen Schmidt's research focuses on social cognition, specifically empathy, racial bias, and how social and environmental feedback can influence self-perception.

Eldar Shafir

William Stewart Tod Professor of Psychology and Public Affairs, Princeton University; Scientific Director and Founder, ideas42

Eldar Shafir's research centers around judgment and decision making and issues related to behavioral economics, with a special interest in poverty and the application of behavioral research to policy.

J. Nicole Shelton

Professor, Department of Psychology, Princeton University

Nicole Shelton's research focuses on understanding prejudice and discrimination from the target's perspective.

Paul Slovic

Professor, Department of Psychology, University of Oregon; President, Decision Research Group

Paul Slovic studies the influence of risk-perception and affect on decisions concerning the management of risk in society, with a special interest in the psychological factors that contribute to apathy toward genocide.

Nancy K. Steblay

Professor, Psychology Department, Augsburg College

Nancy Steblay's research interests are in psychology and law, specifically eyewitness accuracy, pretrial publicity, and inadmissible evidence and judicial instructions to disregard it.

Cass R. Sunstein

Felix Frankfurter Professor of Law, Harvard Law School; Administrator of the White House Office of Information and Regulatory Affairs

Cass Sunstein's research involves the relationship between law and human behavior and particularly focuses on constitutional law, administrative law, environmental law, and law and behavioral economics.

Richard H. Thaler

Ralph and Dorothy Keller Distinguished Service Professor of Behavioral Science and Economics, The University of Chicago Booth School of Business

Richard Thaler studies behavioral economics and finance as well as the psychology of decision making as it influences the conversation between economics and psychology.

Tom Tyler

Professor, Department of Psychology, New York University

Tom Tyler's research is concerned with the dynamics of authority within groups, organizations, and societies and focuses on factors that shape people's motivations when dealing with others in group settings.

Peter Ubel

John O. Blackburn Professor of Marketing, Fuqua School of Business; Professor, Sanford School of Public Policy, Duke University

Peter Ubel's research explores the role of values and preferences in health care decision making, including topics like informed consent, shared decision making, and health care rationing.

Kevin G. Volpp

Professor of Medicine and Health Care Management, The Wharton School, University of Pennsylvania

Kevin Volpp focuses on developing and testing innovative behavioral economics applications to improve patient health behavior and affect provider performance.

Brian Wansink

John Dyson Professor of Marketing and Nutritional Science, Cornell University

Brian Wansink researches food-related consumer behaviors, with a focus on "mindless eating," the study of how microenvironments influence what and how much people eat.

Elke U. Weber

Jerome A. Chazen Professor of International Business; Professor of Psychology and Earth Institute Professor, Columbia University

Elke Weber's research focuses on decision making under uncertainty and on individual differences in risk taking and discounting, specifically in risky financial situations and around environmental decisions.

Andrew K. Woods

Climenko Fellow; Lecturer on Law, Harvard Law School

Andrew Woods's research interests include international human rights and criminal law, with a particular focus on interdisciplinary approaches to law and policy.

David Zionts

Special Advisor to the Legal Adviser, U.S. Department of State

David Zionts's research interests include international law, human rights, and U.S. foreign relations and national security law.

Acknowledgments

The Behavioral Foundations of Public Policy

Introduction

ELDAR SHAFIR

If you look in the dictionary under *policy*, *public policy*, or *social policy*, you find definitions that amount to the following: a system of regulatory measures, laws, principles, funding priorities, guidelines and interventions promulgated by a person, group, or government for the changing, maintenance or creation of living conditions that are conducive to human welfare. Mostly what these measures, laws, principles, and interventions are intended to do is to shape society in desirable ways: to promote behaviors that yield outcomes conducive to human welfare. Successful policy, therefore, must depend on a thorough understanding of human behavior. What motivates and incentivizes people when they snap into action as opposed to procrastinate, obey or disobey the law, understand or misunderstand, act or fail to act on their intentions, care or do not care, attend or get distracted? How do they perceive their decisions and the options at their disposal? How do they think about what others are doing? These are all questions that must be addressed for the design and implementation of policies to prove successful.

In light of the centrality of behavioral assumptions to policy, it is remarkable how small a role the attempt to understand human behavior has played in policy circles, as well as in the social sciences more generally. It is particularly remarkable because, as we have now come to understand, much of our intuition about human behavior fails to predict what people will do. And policies based on bad intuitive psychology are less likely to succeed and can often prove hurtful. As the economist John Maurice Clark pointed out nearly a century ago, if the policy maker does not seriously study psychology, "he will not thereby avoid psychology. Rather, he will force himself to make his own, and it will be bad psychology" (*Journal of Political Economy*, 1918).

Bad psychology comes in many forms. A naive understanding of incentives, for example, might suggest that paying people some small amount (rather than nothing) to perform a societally desirable act could only increase instances of that act; instead, it turns out that the loss of the "psychic" benefit of having been a good citizen (which is largely neutralized by the monetary remuneration) can, in fact, reduce take-up. Alternatively, presenting lineups (where suspects are observed concurrently) versus show-ups (where they are seen one at a time) may appear normatively indistinguishable, but we now know that the former leads to more false identifications than the latter. Similarly, having workers opt out of, rather than opt into, retirement savings accounts, looks like an immaterial nuance, except that the former, for predictable reasons and for what amounts to very little cost, generates many more happy retirees than the latter.

A careful consideration of the role of psychology in public policy took many years to develop even after Clark's warning about the dangers of bad psychological assumptions. An important turning point was the behavioral critique of the economic assumptions underlying individual decision making begun by cognitive and social psychologists in the 1970s. This was eventually reinforced by the economic profession's gradual, even if reluctant, acceptance of the behavioral critique and led to increased research applying behavioral insights to studies of choice and judgment in everyday life. Now, almost a half century after the emergence of the modern critique, the behavioral perspective occupies a respectable and increasingly popular niche in many graduate programs in economics, business, law, policy, and the social sciences more generally. And thus we have arrived at a point where it is only natural to explore how best to incorporate elements of the behavioral perspective into policy thinking.

The behavioral findings provide an alternative view of the human agent. Many aspects of decision making that the normative analysis assumes do not matter (such as how the options are described, as long as the same information is given) prove highly consequential behaviorally, and other factors that are normatively assumed to be of great importance (such as whether an intervention will help save 1,000 birds or 10,000 birds) are, instead, intuitively largely ignored. At the most general level, a couple of deep lessons have emerged that are of great potential relevance to policy makers: the relevance of context and the unavoidability of construal.

Human behavior tends to be heavily context dependent. One of the major lessons of modern psychological research is the impressive power that the situation exerts, along with a persistent tendency on our part to underestimate this power relative to the presumed influence of personal intentions and traits. In his classic obedience studies, for example, Milgram (1974) demonstrated the ability of decidedly mild situational factors to trigger behaviors on the part of regular citizens, such as the administration of presumed electric shocks to innocent others, that were unfathomable to naive introspection. Along similar lines, Darley and Batson (1973) showed how seminary students on their way to deliver a sermon on the parable of the Good Samaritan were willing to walk right past an ostensibly injured man, slumped coughing and groaning, simply because they were running late. Minor contextual features were shown in these cases to override people's professed intentions and their deeply held convictions. To the extent that such minor contextual features are able to transcend education, personality, and intention, policy makers appear to have powers of influence that they underappreciate, may unintentionally misuse, and could, given some behavioral insight, employ better.

The second lesson, which is fundamental to the cognitive sciences in general, concerns the role of "construal" in mental life. People do not respond to objective experience; rather, stimuli are mentally construed, interpreted, and understood (or misunderstood). While this claim risks sounding deep, it is actually trivial, but with profound consequences: behavior is directed not toward actual states of the world but toward mental representations of those states, and those representations do not bear a one-to-one correspondence with the states they represent. In fact, the representations we construct may not even constitute faithful renditions of actual circumstances. Our visual experience, for example, is the product of complex processes that take raw visual input (say, a trapezoid when we look at a window from the side) and use contextual cues to represent what is "really there" (a perfectly rectangular window). Anytime those cues are misleading, we end up with a false representation, as in the case of well-known optical illusions. How we interpret attitudes and emotions is similarly a matter of construal. And, as it turns out, so is our representation of many objects of judgment and preference. We can only decide between things as they are represented in the three-pound machine that we carry behind the eyes and between the ears. And those representations are the outcome of mental processes that, to some extent at least, have a life of their own.

For policy makers all this should be of the utmost importance. Policies' success depends on human behavior. And behavior is determined not simply by what is available, but by what people know, perceive, understand, attend to, and want. Thus, well-intentioned interventions can fail because of the way they are construed by the targeted group. And the difference between success and failure can sometimes boil down to a relatively benign and normatively immaterial change in presentation and construal, rather than a complex and costly rearrangement of the available alternatives.

About fifteen years ago, we began a joint formal program of training in "psychology for policy" between the psychology department and the Woodrow Wilson School of Public and International Affairs at Princeton University. The endeavor was new at the time, and the results of the initiative were not entirely predictable. What were some of the more important policy questions to which a behavioral analysis could significantly contribute? Where in policy did misguided behavioral assumptions figure most prominently, where were they of lesser importance, and what exactly were they anyway? How was one to go about researching and communicating all this? And would it make a difference? As often happens when ideas gather momentum, we were not alone. An increasingly talented and interdisciplinary group of scholars had grown interested in research along similar lines and in issues of both behavioral and policy significance.

The present volume presents some of the more impressive outcomes of this important work, as conceived and summarized by many of the leading scholars in the field. The wide array of topics covered here should appeal to students of human behavior interested in real-world applications. More importantly, the chapters in this volume were prepared with an eye toward a sophisticated audience with no behavioral training. The application of experimental findings and concepts emanating from behavioral research to the design and implementation of policy—call it "behavioral policy"—is an exciting and rapidly expanding new area of research and study. The present collection is intended to expose policy makers, practitioners, government officials, business leaders, legal, ethical, and health professionals, as well as students interested in societal, domestic, and international challenges, to a perspective that can shed new light. Greater insight into human behavior, the authors in this volume agree, can prove helpful, both in making sense of what are otherwise persistent puzzles, as well as in generating novel ideas and effective solutions.

The contributions to this collection tend to be highly interdisciplinary and thus hard to compartmentalize. Nonetheless, the sheer amount of material presented in this volume warrants some minimal organization in the hopes of facilitating the reader's task. Chapters are divided by general topics but are

otherwise independent and can be read in any order; occasional cross-references occur when the materials of separate chapters are especially complementary. The aim of each chapter is to provide the reader with an overview of how research in the behavioral sciences might influence our understanding and the conduct of good policy in a particular domain. Ultimately, we hope the reader will come to see the foundational role that behavioral assumptions must come to play in shaping the successful design and implementation of policy.

The Chapters

The early chapters focus on behavioral issues that arise in the conduct of our social and political lives. They focus on policy-relevant topics ranging from the nature of intuitive social judgment and "automatic" social perceptions to the valuation of social belonging and concerns with identity, justice, and fairness; problems ranging from discrimination in the work place to the numbing that comes with hearing about mass atrocities.

A common thread running through these contributions is that the empirical findings are often in tension both with normative assumptions as well as with common intuition. As a result, they have far-reaching consequences for how we think about policy design and implementation. We tend to think, for example, that people's behavior largely mirrors their beliefs and that their choices are typically about tangible, value-maximizing outcomes. Thus, we might assume, those who are not prejudiced will typically not exhibit prejudiced judgment, and if voting is unlikely to have a tangible impact, people will not bother to vote. Similarly, the intuition goes, negotiators whose persistent biases lead to impasse will learn to overcome them, and managers whose unintended discriminatory practices hurt their business will learn to avoid discriminating.

In contrast to all that, as the chapters below illustrate, people care a lot about intangibles, they exhibit persistent biases in social perception, and they lack introspective access to the biases and the motivations that often are in tension with their better judgment. As a result, people often fail to recognize the discrepancies between their beliefs and their actions, which, rather than resolving themselves in the long run, often end up playing a big role in exacerbating long-standing political and social tensions.

Prejudice and Discrimination

In the opening chapter, on implicit prejudice, **Curtis Hardin** and **Mahzarin Banaji** argue that our views of prejudice and discrimination are based on outdated notions, with important policy implications. Rather than arising from ignorance and hatred, which would be best addressed by changing the hearts and minds of individuals, prejudice and stereotyping, according to these authors, emerge from cognitively salient structures in our social milieu and do not necessarily involve personal animus, hostility, or even awareness. Rather, prejudice is often "implicit"—that is, unwitting, unintentional, and uncontrollable—even among the most well intentioned. At the same time, these authors suggest, research shows that implicit prejudice can be reduced through sensible changes in the social environment.

The social environment figures prominently in **Nicole Shelton**, **Jennifer Richeson**, and **John Dovidio**'s chapter on intergroup biases in interracial interactions. The goal of this chapter is to explore how racial bias can influence affective, cognitive, and behavioral outcomes during interracial interactions, especially among those who do not harbor explicitly racist attitudes. This question is examined in a variety of contexts, including students sharing dorm rooms on university campuses and interactions between White physicians and racial minority patients in health care settings. A central message that emerges from these interactions is that bias is expressed in subtle ways: as strained relationships between roommates, as less effective interactions between physicians and patients, and as lower levels of rapport in employment interviews. In each of these cases, there is rarely an obvious act of blatant discrimination. Instead, the complex and often subtle nature of contemporary intergroup bias, for which traditional policies designed to respond to overt discrimination are ill suited, can have widespread impacts on intergroup interactions, often with different consequences for members of different racial and ethnic groups. Shelton et al. conclude with a review of common practices and interventions that policy makers could use to maximize the benefits of diversity across policy-relevant settings.

In their chapter on gender bias, **Susan Fiske** and **Linda Krieger** consider the legal ramifications of unexamined gender discrimination, particularly as it plays out in employment contexts. They review recent behavioral and neuroscience research that challenges the rational-actor assumption underlying much of the debate over discrimination law and policy. Decision makers, according to Fiske and Krieger, cannot always make optimal employment decisions, because, even when they consciously support equal opportunity norms, subtle, unexamined forms of gender bias may prevent them from accurately perceiving decision-relevant information, or from optimally using it to make employment decisions. Managers may explicitly endorse equal opportunity, but unexamined prejudices

might nevertheless derail their choices. Fiske and Krieger consider the kinds of initiatives that organizations might undertake in an attempt to reduce the levels of workplace discrimination caused by unexamined, subtle bias. They also advocate for policy initiatives, including a mandatory information disclosure approach to equal opportunity employment policy, which they suggest might help squeeze discrimination out of labor markets in ways that circumvent the need to identify individual instances of discriminatory decision making.

Social Interactions

In his chapter on the psychology of cooperation, **Tom Tyler** argues that, while incentives and sanctions matter, standard normative approaches place too much emphasis on issues of material gains and losses. Tyler analyzes laboratory and field studies that illustrate several types of social motivations—attitudes, values, personal identity, procedural justice, and motive-based trust—that have a strong influence on behaviors in social settings. Tyler focuses on the problem of motivating cooperative behavior and suggests that policy makers have a great deal to gain from expanding their conception of human nature and recognizing the importance of social motivations in shaping people's behavior in groups and organizations.

Along related lines, **Todd Rogers**, **Craig Fox**, and **Alan Gerber** propose an alternative conceptualization for why people vote. Rather than the standard self-interested view, which cannot explain the decision to vote given the miniscule probability that one's vote will affect the outcome, the authors propose to think of voting as a "dynamic social expression." Voting, according to this perspective, is the outcome of a dynamic constellation of events that extend over time; it is an inherently social act, and it is ultimately an expression of one's identity. Among other things, Rogers, Fox, and Gerber describe recent experimental field research into effective get-out-the-vote campaigns, thereby linking the question of why people vote to an array of behavioral research—including social and cognitive psychology and behavioral economics—that has not been systematically linked to voting behavior in the past.

In his chapter on disagreement, **Lee Ross** considers several cognitive and motivational processes and the role they play in adding hostility and distrust to policy disagreements, and how they serve as barriers to dispute resolution. Among other constructs, he considers the *reactive devaluation* of proposals put forth by the other side and the role of *naive realism*, the conviction that one sees things objectively and clearly, which tends to add rancor to disagreement

insofar as it creates the expectation that other reasonable and honest people ought to share one's views. (This perspective was well captured by comedian George Carlin's observation about driving: "Ever notice that anyone going slower than you is an idiot and anyone going faster is a maniac?") Informed by the foregoing analysis, Ross then considers several behaviorally informed strategies for overcoming barriers to agreement.

Finally, in their chapter on psychic numbing, **Paul Slovic**, **David Zionts**, **Andrew Woods**, **Ryan Goodman**, and **Derek Jinks** ask why people repeatedly fail to react to genocide and other mass-scale human atrocities. It is not, they argue, because people are insensitive to the suffering of their fellow human beings, or even that most only care about identifiable victims of similar skin color who live nearby. Rather, they suggest, a fundamental problem lies in people's incapacity to experience commensurate *affect*, the positive and negative feelings that combine with reasoned analysis to guide action. Left to its own devices, moral intuition appears to respond more to individual stories that are easier to imagine than to statistics of mass atrocities, which fail to spark commensurate affect and motivate appropriate action. Even when we know genocide is real, we do not "feel" that reality. The authors explore some behaviorally informed ways that might make genocide "feel real," but they are ultimately led to the conclusion that we cannot rely on intuitive reactions and must instead commit ourselves to institutional, legal, and political responses that are less susceptible to psychic numbing and more heavily based upon reasoned analysis of our moral obligations.

The Justice System

The rational agent model has figured prominently in many areas of the law. At the same time, much of what comes under the law depends on the impulses, intuitions, judgments, sense of confidence, emotional reactions, and everyday understandings of regular citizens when they act as witnesses, jurors, colleagues, employers, employees, and so forth. And because the legal system is heavily in the business of constructing rules and procedures, there is much room to think about how these can be better shaped by a nuanced understanding of human capabilities, proclivities, and limitations.

In their chapter on eyewitness identification and the legal system, **Nancy Steblay** and **Elizabeth Loftus** focus on issues of eyewitness memory, such as the fact that faulty eyewitness memory has been implicated in a majority of (mostly DNA-based) exonerations. They review the main lessons from the science of eyewit-

ness memory and consider their implications for improving the legal system whenever eyewitnesses are playing a crucial role. They provide a primer on the essential memory principles underlying eyewitness performance, including the fact that this experience is not just a memory phenomenon, but that it also reflects social forces, including, for example, subtle and unintentional verbal and nonverbal communications from others. What emerges is the potential for memory to be contaminated, distorted, and yet reported with great confidence, which proves of great relevance to a legal system that depends on and believes in eyewitness veracity and in which many people become criminal defendants on the basis of eyewitness identification. Steblay and Loftus describe the ongoing research effort around the topic of eyewitness testimony, the changes in legal policy spurred by the collaboration between behavioral scientists and those in the legal field, and the challenges that persist in the application of memory research to public policy.

In "False Convictions," **Phoebe Ellsworth** and **Sam Gross** extend the analysis of the rate and persistence of false convictions to a variety of psychological, social, and institutional factors beyond eyewitness identification. They first highlight the inherent difficulty of detecting false convictions, where the only ones we know of (and even there we could be wrong) are exonerations: those rare cases in which a convicted criminal defendant—typically in the most serious of cases, the only ones to receive sufficient attention and resources—is able to prove his innocence after the fact. Ellsworth and Gross consider the social and institutional context that characterizes criminal investigation and adjudication under the adversarial system. They analyze the proclivity of the process to give rise to worrisome behavioral phenomena, including confirmation biases, eyewitness misidentification, false confessions, fraud and error on the part of forensic analysts, perjury by jailhouse informants and other witnesses, misconduct by police and prosecutors, and incompetent representation by criminal defense attorneys. Ellsworth and Gross describe the relevant work by social scientists and legal researchers and consider some areas for future policy enhancement.

In a chapter focusing on behavioral reactions to wrongdoing, **John Darley** and **Adam Alter** explore the nature and consequences of potential gaps between legal codes and community sentiments regarding punishment, retribution, and deterrence. They first review research on people's perceptions of wrongful actions and the punishments those actions deserve. They conclude that people are driven by emotionally tinged reactions of moral outrage and that their punishment decisions are largely based on what they intuitively believe the offender justly deserves. They then consider conventional approaches to dealing with crime, punishment, and deterrence and conclude that in light of what we know about human cognition and behavior, those approaches are largely ineffective. For example, whereas our penal system focuses heavily on sentence duration, sentence duration is generally an ineffective deterrent, as compared, for example, to salient surveillance mechanisms. Darley and Alter consider relevant policy implications, while keeping in mind that citizens' intuitive perceptions of justice will place limits on the types of societal punishment practices that will be perceived as fair and that legal codes that clash with those moral sensibilities can cause citizens to lose respect for the law.

Bias and Competence

Of great relevance to policy are the circumstances in which people exhibit systematic bias or fail to weigh appropriately the factors that matter most to a decision. In other circumstances, people may perform the requisite tasks exceedingly well. This contrast is heightened by the fact that it is often hard for people to anticipate when they might expect bias as opposed to remarkable judgmental acuity. Things that ought not matter from a normative perspective often do, and things that ought to matter often fail to have an impact. The chapters in this section are motivated by the assumption that greater awareness and the proper anticipation of bias and other behavioral limitations may help devise more effective policies.

In their chapter on claims and denials of bias, **Emily Pronin** and **Kathleen Schmidt** explore the far-reaching policy implications of people's perception that their own judgments are relatively free of bias whereas others' judgments are susceptible to it. People's tendency to be blind to their own biases while exaggerating those of others can lead to a range of problems, among which are social conflict, breakdown of negotiations, corruption, and discrimination. Pronin and Schmidt examine the central behavioral underpinnings of this "bias blind spot" and consider potential solutions, including increased awareness, education, and psychologically savvy disclosure requirements, with an emphasis on how to make those solutions psychologically informed and thus more effective.

In "Questions of Competence: The Duty to Inform and the Limits to Choice," **Baruch Fischhoff** and **Sara Eggers** discuss the nature of assumptions about people's competence that figure, often implicitly, in a wide range of regulatory and policy domains. For example, product disclosure requirements reflect beliefs about people's ability to recruit and comprehend the relevant information, and policies governing

living wills reflect assumptions about our ability to anticipate the relevant circumstances. When competence is underestimated, they argue, people's freedom to make their own decisions may be needlessly curtailed, whereas when competence is overestimated, they may be denied important protections. The chapter considers several applications to risk-related decisions in U.S. policy contexts (including drugs, pathogens, and contaminants) in its attempt to offer a general approach to assessing and, where possible, improving individuals' competence to make the requisite decisions.

In his chapter on misfearing and cost-benefit analysis, **Cass Sunstein** argues in support of cost-benefit analysis as a way of counteracting the problem of misfearing, that is, people's tendency to misperceive risks. Whereas cost-benefit analysis is often justified on conventional economic grounds, as a way of preventing inefficiency, Sunstein argues for it on grounds associated with cognitive and social psychology, including concepts such as availability and salience, informational and reputational cascades, intense emotional reactions, motivated reasoning, and causal misattribution, all of which can lead people to be afraid of fairly trivial risks and neglectful of serious ones. Such misfearing, Sunstein suggests, plays a significant role in public policy because of the power of self-interested private groups and ordinary political dynamics. And when misallocations of public resources result from misfearing and associated problems, cost-benefit analysis can operate as a corrective, a way of ensuring better priority setting and of overcoming behavioral obstacles to desirable regulation.

Behavioral Economics and Finance

Behavioral research in economics and finance has explored the systematic ways in which people's preferences are in tension with standard assumptions underlying the classical theories of choice. Among other things, people tend to focus on perceived departures from the status quo rather than on final assets, they exhibit unstable discount rates, and they tend to be loss averse—the dread of losses is greater than the savoring of equivalent gains. Intangibles such as fairness and inertia matter a lot, and decisions are often made "locally," with much reliance on features that loom large at the moment, often at the expense of long-term objective outcomes. All this puts a greater burden on policy design, since minor and normatively inconsequential changes can make the difference between policies that succeed and those that fail.

This type of analysis, in the context of retirement saving plans, is illustrated by **Shlomo Benartzi**, **Ehud Peleg**, and **Richard Thaler**, who apply behavioral principles to the study of the choices made by employees saving for retirement. Exploring notions ranging from decision inertia and nominal loss aversion to discounting and the synchronization of saving increases with pay raises, they show how supposedly minor details in the architecture of retirement plans can have dramatic effects on investment decisions and savings rates. More generally, they suggest, such insights into the architecture of decision have the potential to help people make better decisions, a theme Thaler returns to in a chapter with Balz and Sunstein later in the book.

Applying a behavioral economic analysis to employment law, **Christine Jolls** considers the lessons of behavioral analysis for legal requirements and rules that govern employer-employee relationships, ranging from wage payment and pension regulation to minimum wages, mandated health insurance, workplace leave, and discrimination laws. The effects of employment law, Jolls argues, turn in significant part on people's behavior in employment settings, which can be illuminated by consideration of bounded willpower, bounded self-interest, and bounded rationality. Thus, errors in intuitive judgment have implications for employment discrimination law, and different rules may prove more effective in encouraging retirement saving by individuals with bounded willpower. Furthermore, a "fairness dynamic"—one in which employers choose to pay employees more than the minimum they would accept and employees respond by working harder than they otherwise would—has implications for minimum wage regulation. The employment relationship, Jolls concludes, is one of life's most important relationships, and it can greatly benefit from a behavioral economic perspective.

In their chapter on decision making and policy in contexts of poverty, **Sendhil Mullainathan** and **Eldar Shafir** present a behaviorally motivated framework for understanding the decisions of the poor. Motivated by empirical insights on judgment and decision making that are supplemented by lessons from social and cognitive psychology, they ask how we might explain behaviors in poverty and how might similar behaviors have different consequences when people are poor. They conclude with recent work in which poverty is viewed as a context that creates unique challenges for the human psyche, above and beyond budgetary woes. Poverty itself, according to Mullainathan and Shafir, generates specific psychological responses that are endemic to functioning with little slack and with constant vigilance and that can lead to distraction, miscalculation, and depletion. This, they propose, suggests new approaches to policy making that are focused on programs that foster stability and give people the financial and psychic steadiness needed to build more robust economic lives.

Behavior Change

Subtle changes in the context of decision can have a significant, and normatively unexpected, impact on the course of action taken. And this has important implications for the familiar tension between intention and action. In the face of contextual obstacles (which can range from transportation to shame to forgetfulness), people can fail to act even when they have a strong intention to do otherwise. In contrast, when the context is designed to facilitate certain actions, those actions might be taken even when resolve is not terribly high. Contextual cues and interventions, incentives, and decision aids, sometimes quite subtle and limited in scope, can have substantial impact even in contexts where preferences otherwise appear clear and strong. This, of course, only increases the onus on policy makers to construct and implement policies that are behaviorally insightful and thus more likely to have the desired effects. It also suggests that at times the passage from a policy that is not working to one that does may require different, and perhaps more nuanced (and affordable), changes, ones that address how people construe a problem and what that construal leads them to do.

In their chapter on the psychological levers of behavior change, **Dale Miller** and **Deborah Prentice** focus on circumstances in which policy makers wish to help people change their behavior in ways that align with these people's own long-term interests and stated wishes. Miller and Prentice analyze the capacity of various interventions to move people toward desirable behavior and away from undesirable behavior, with a special emphasis on the psychological constructs and processes that produce behavior change. Among other things, they illustrate ways in which economic and psychological incentives can combine in complex ways, producing counterintuitive effects from economic taxes and subsidies. They outline how efforts to change behavior must begin with a careful analysis of the motivational dynamics bearing on the status quo and the levers that can be used to change them.

In his chapter, "Turning Mindless Eating into Healthy Eating," **Brian Wansink** considers some of the basic processes behind a variety of environmental factors that influence food consumption. Package size, plate shape, lighting, socializing, and the visibility, variety, size, and accessibility of food are only some of the environmental factors that influence the volume of food consumed and are considered likely to have contributed to an ever-widening obesity problem in many places. Understanding these drivers of consumption volume has immediate implications for nutrition education and consumer welfare, but education and increasing awareness are unlikely to be the

solutions, Wansink argues, because the effects occur at perceptual levels of which we are not aware. Instead, he lists some behaviorally informed principles that academics, industry, and government can use when partnering to make tangible health-related changes in the lives of individuals.

In "A Social Psychological Approach to Educational Intervention," **Julio Garcia** and **Geoffrey Cohen** focus on the psychological causes of academic underperformance, particularly the racial achievement gap observed in American schools. Among others, they describe psychological interventions that focus on the presence of an "identity threat" and that when systematically applied have been found to close the achievement gap. At the heart of their analysis is the notion of the classroom as a tension system in which various factors, both structural and psychological, interact to produce an environment that elicits a set of attitudes, behaviors, and performance. By heightening the impact of factors facilitating performance or lessening the impact of factors that impede it, interventions can alter students' psychological environments. This analysis leads Garcia and Cohen to conclude that well-timed interventions targeting important psychological processes can produce effects on performance that appear disproportionately large. Throughout, they discuss the implications for social policy that follow from their approach.

Improving Decisions

Systematic tendencies, ranging from an inadequate weighing of likelihoods to excessive discounting of the future to an inability to simulate future feelings, can all interfere with the making of optimal decisions. Furthermore, limited mental resources and attention have important implications for people's abilities to budget, save, invest in mitigation against natural disasters, or bother to develop a long-term collective perspective. What repeatedly emerges as important is not sheer human ability, which can be impressive, but the fact that intuition, attention, and understanding can be tapped into in ways that are less or more likely to succeed. The contributions that follow consider behaviorally informed ways in which policy makers might help people reach better decisions, individually and collectively.

Going beyond issues of mere comprehension, **Peter Ubel** considers the use of medical decision aids to improve people's "preference sensitive decisions," where the decision maker, when left to her own devices, might not make the right choice. In particular, what is envisioned is a neutral party to help the patient make a decision consistent with her underlying goals and preferences. Experts on decision aids, Ubel argues,

have typically assumed that if you give decision makers full information and the freedom to choose, they will experience reduced conflict and higher satisfaction and will make decisions that reflect their true preferences. Instead, he suggests, decision counselors need to go beyond increased comprehension and conflict reduction, in light of much evidence showing that people who comprehend their options nevertheless can make bad decisions, and that good decisions can still leave people deeply conflicted. Toward that end, Ubel evaluates the strengths and weaknesses of several criteria by which to determine whether a structured decision aid has helped people make good preference-sensitive decisions.

In their chapter, "Using Decision Errors to Help People Help Themselves," **George Loewenstein**, **Leslie John**, and **Kevin Volpp** argue that having identified a variety of systematic decision errors, behavioral researchers are in a good position to provide policy solutions that make use of those same errors to people's benefit. They show how a wide range of decision phenomena that are typically viewed as errors—including the status quo and default bias, loss aversion, overoptimism, and nonlinear probability weighting, among others—can be exploited to help people accomplish their goals, ranging from saving money and losing weight to drug adherence and charitable giving. They also consider whether such errors could be exploited to deal with broader societal problems such as global warming. There are, according to Loewenstein, John, and Volpp, many economic entities that exploit consumers' mistakes. Instead of leaving consumers to fend for themselves, we ought to harness the same errors that are regularly used to exploit them to instead help make people better off.

In her chapter exploring the insights of behavioral decision research for better environmental decisions, **Elke Weber** starts by outlining the logic of environmental policy decisions, which typically include social and economic dimensions, considerations of fairness or equity and considerable uncertainty involving intertemporal tradeoffs and which require foresight, patience, and persuasion. Because environmental goods like clean air, drinkable water, and species diversity are common-pool resources, rational analysis essentially prescribes shortsighted behaviors even if more longsighted and cooperative solutions are socially desirable. Informed by social cognition and behavioral decision research, Weber argues that insights into unconscious and social inferential and decision processes, as well as into people's limitations in attention, memory, and information processing, can help guide the design of more promising environmental policies. Behaviorally informed considerations, she argues, suggest that people might be induced to act in more

collective ways that increase their own long-term benefits, as long as we are able to shape their decision environment in ways that facilitate environmentally sustainable behaviors.

In their chapter on overcoming decision biases to reduce losses from natural catastrophes, **Howard Kunreuther**, **Robert Meyer**, and **Erwann Michel-Kerjan** describe the recent trend of escalating losses from natural hazards. They attribute this to an interplay of economic and behavioral factors: increased levels of assets placed in harm's way often without adequate investments in mitigation, along with a tendency to underattend to low-probability, high-consequence risks and to underappreciate the benefits of long-term investments in protection. The result is an accelerating spiral of risk taking, where the rate of economic development in high-risk areas outpaces investment in technologies intended to protect those developments. Kunreuther, Meyer, and Michel-Kerjan consider some of the behavioral drivers of this mismatch, and how taking these into account might help devise instruments (such as long-term insurance policies coupled with home improvement loans to induce investment in cost-effective mitigation measures) that can help reduce losses from future natural disasters.

Decision Contexts

The concluding three chapters explore several important features of contextual design—defaults, choice architecture more generally, and behaviorally informed regulation—all of which, it is argued, can aid in the implementation of improved policies. In "Decisions by Default," **Eric Johnson** and **Daniel Goldstein** draw on a variety of policy domains to illustrate the power of defaults and then explore some of the psychological mechanisms that underlie these effects. From insurance and organ-donation decisions to retirement savings and internet privacy settings, changing a no-action default can be highly effective compared to economic incentives or extensive educational or persuasion campaigns designed to influence people to make active decisions. Guided by the realization that each kind of default has costs and benefits and by considerations of ethics and effectiveness, Johnson and Goldstein discuss the importance to policy makers of understanding defaults and suggest conditions when different kinds of default arrangements—forced choice; mass defaults; random, smart, or personalized defaults—might be advisable.

In their chapter on choice architecture, **Richard Thaler**, **Cass Sunstein**, and **John Balz** consider decision makers, who—like all of us, if you believe the behavioral findings—function in an environment where many features, noticed and unnoticed, can influence

the decisions that they make. Those who shape the decision environment, in this case the policy makers, are the "choice architects." Thaler, Sunstein, and Balz analyze some of the tools that are available to choice architects, such as creating defaults, expecting errors, understanding mappings, giving feedback, structuring complex choices, and creating incentives. Their goal is to show how choice architecture can be used to help nudge people to make better choices (often as judged by themselves) without forcing the intended outcomes upon anyone, a philosophy they call libertarian paternalism.

Finally, looking at behaviorally informed regulation, **Michael Barr**, **Sendhil Mullainathan**, and **Eldar Shafir** propose a regulatory framework based on insights from behavioral economics and industrial organization in which outcomes are an equilibrium interaction between individuals with specific psychologies and firms that respond to those psychologies within specific market contexts (in contrast to the classic model, which assumes an interaction between rational choice and market competition). The introduction of a richer psychology, Barr, Mullainathan, and Shafir propose, complicates the impact of competition. It suggests that firms compete based on how consumers respond, and competitive outcomes may not always align with increased consumer welfare. Regulation must then address failures in this equilibrium. For example, in some contexts market participants will seek to overcome common human failings (as for example, with undersaving), whereas in other contexts market participants will seek to exploit them (as with overborrowing). Barr et al. discuss specific applications and illustrate, among other things, how a behaviorally informed regulatory analysis could im-prove policy makers' understanding of the costs and benefits of specific policies.

Commentaries

The volume concludes with a series of commentaries from scholars in four disciplines—philosophy, economics, medicine, and law. These scholars' main lines of research lie outside the behavioral arena, but they all have had a longstanding interest in behavioral applications and took it upon themselves to comment on issues raised in this volume, particularly as they interact with their own disciplinary ways of thinking. **William Congdon** considers some of the ways in which the behavioral perspective can inform economic policy; **Donald Redelmeier** looks at the ways in which behavioral insights might inform health care policy; **Paul Brest** focuses his attention on issues surrounding the potential debiasing of policy makers and lawmakers; and **Judith Lichtenberg** aims a philosophical lens at issues of paternalism, manipulation, and the extent to which behaviorally informed policy making may be good for people.

In the chapters that follow, more than fifty scholars will tell you about a rich body of research conducted over the past three to four decades that has changed the way we understand people. They will consider several implications of the research findings, and they will suggest many ways in which our new understanding, this new view of the human agent, might help design and implement better public policy. We hope that you find this exposition of the behavioral foundations of policy productive and illuminating and that you will use it to create new policies that further improve human welfare.

PREJUDICE AND DISCRIMINATION

The Nature of Implicit Prejudice

Implications for Personal and Public Policy

CURTIS D. HARDIN

MAHZARIN R. BANAJI

Some fifty years ago in Arkansas, nine black students initiated a social experiment with help from family, friends, and armed National Guards. Their successful attempt to desegregate Little Rock's Central High School following the decision in *Brown v. Board of Education* is among the most momentous events in America's history, leaving no doubt about its historic importance and the significance of its impact on public policy. Nevertheless, as many have noted, even at the beginning of the twenty-first century, a blatant de facto segregation in living and learning persists and in some circumstances has intensified (e.g., Orfield, 2001). The American experiment in desegregation is a reminder that public policies, however noble in intent, may not realize their aspirations if they do not include an understanding of human nature and culture. In other words, they cannot succeed if they are not founded on relevant scientific evidence, which reveals the nature of the problem, the likely outcomes, and how social transformation can best be imagined. As an example of the importance of basing policy in science, there is the research of Robert Putnam showing the unsavory result that ethnic diversity may actually increase social distrust. As the ethnic diversity by zip code increases, so does mistrust of one's neighbors, even same-ethnicity neighbors (Putnam, 2007). The naive optimism that diversity will succeed in the absence of a clear understanding of the dynamics of social dominance and intergroup relations is challenged by these and other similar revelations (e.g., Shelton, Richeson, and Dovidio, this volume). Hence, even well-intentioned public policies are unlikely to yield positive outcomes unless they are grounded in the best thinking available about how people actually think and behave. Sadly, this has not been the case, both because policy makers are not sufficiently respectful of the importance of science as the guide to social issues and because academic scientists resist imagining the policy implications of their evidence.

In this chapter, we address the topics of stereotyping and prejudice, staying firmly within the bounds of what science has demonstrated. However, in keeping with the mission of this book, we spell out what we see to be some obvious, and also some less obvious, tentacles to questions of public policy. We posed the following questions to ourselves: What are the broad lessons learned that have changed our understanding of human nature and social relations in recent decades? In what way does the new view run counter to long-held assumptions? How should policy involving intergroup relations proceed in light of these discoveries? And, can we speak about "personal policies" that may emerge from the education of individuals about the constraints and flexibility of their own minds while also considering the notion of policy in the usual "public" sense? Our contention is that personal and public policy discussions regarding prejudice and discrimination are too often based on an outdated notion of the nature of prejudice. Most continue to view prejudice as it was formulated generations ago: negative attitudes about social groups and their members rooted in ignorance and perpetuated by individuals motivated by animus and hatred. The primary implication of the old view was that prejudice is best addressed by changing the hearts and minds of individuals, for good-hearted people will think well of others and behave accordingly. However, research in recent years demonstrates that the old view of prejudice is incomplete, even dangerously so. Staying with it would lead to policy choices that might be ineffectual, or worse. Staying with it would be akin to ignoring the evidence on smoking and cancer.

How has the scientific understanding of prejudice changed? In short, we now know that the operation of prejudice and stereotyping in social judgment and behavior does not require personal animus, hostility, or even awareness. In fact, prejudice is often "implicit"—that is, unwitting, unintentional, and

uncontrollable—even among the most well-intentioned people (for a review, see Dovidio and Gaertner, 2004). Moreover, although the discovery of implicit prejudice initially brought with it an assumption that it might be unavoidable (e.g., Bargh, 1999; Devine, 1989; Dovidio et al., 1997), research demonstrates that, although it remains stubbornly immune to individual efforts to wish it away, it can be reduced and even reversed within specific social situations through sensible changes in the social environment (e.g., Lowery, Hardin, and Sinclair, 2001; Rudman, Ashmore, and Gary, 2001). In sum, in addition to the real problems that malicious "bad apples" pose for social policy, research demonstrates that prejudice also lives and thrives in the banal workings of normal, everyday human thought and activity. In fact, an overemphasis on the bad apples may well be detrimental to considerations of policy because it assumes the problem of prejudice to be that of the few rather than that of the many (Banaji, Bazerman, and Chugh, 2003).

We believe that the new understanding of prejudice that has evolved over the past three decades invites a transformation of the public debate regarding how the problem of prejudice may be productively addressed. Hence, this chapter will review the research that has so dramatically changed the contemporary understanding of the nature of prejudice, with an emphasis on research demonstrating (a) the existence of implicit prejudice, (b) the ubiquity of implicit prejudice and its consequences, (c) principles by which the operation of implicit prejudice may be influenced, and (d) the policy changes implied by a recognition of what the mind contains and is capable of. In so doing, we argue that although implicit prejudice has disturbing consequences for social judgment and behavior, potential solutions may arise in part from a reconceptualization of prejudice—less as a property of malicious individuals and more as a property of the architecture of cognition and known mechanisms of social learning and social relations.

The Nature of Implicit Prejudice

The discovery that prejudice can operate unwittingly, unintentionally, and unavoidably emerged from several related developments in psychology, sociology, economics, and political science. Most politically salient was the persistence of social, economic, and health-related racial discrimination despite an increasing unwillingness, during the late-twentieth century, of Americans to consciously endorse "explicit" racist attitudes (e.g., Bobo, 2001; Dovidio, 2001; Sniderman and Carmines, 1997). Although

the observation of dissociations between explicit intergroup attitudes and intergroup discrimination was hardly unprecedented (e.g., Allport, 1958; La Pierre, 1934), it was met with an increasing interest in assessing political attitudes unobtrusively, either to circumvent the role of social desirability in attitude expression (e.g., Crosby, Bromley, and Saxe, 1980; Fazio et al., 1995; Word, Zanna, and Cooper, 1974), or to address the possibility that the psychology of prejudice in the United States had evolved into more sublimated, symbolic, or otherwise less deliberately hostile forms (e.g., Dovidio and Gaertner, 2004; Jackman, 1994; Sears and Henry, 2005). Equally important, developments within the information-processing paradigm of psychology made the study of implicit cognition—including automatic, implicit prejudice—both newly possible and theoretically coherent (e.g., Banaji and Greenwald, 1994; Bargh, 1999; Greenwald and Banaji, 1995). Finally, the social-psychological interest in implicit prejudice resonated with a broader interdisciplinary appreciation across the brain sciences of the variety, sophistication, and richness of information processing that occurs outside the window of conscious deliberation, indicating, among many other things, that prejudice is hardly the only kind of thinking largely implicit in nature (e.g., French and Cleeremans, 2002).

The Discovery of Implicit Prejudice

The discovery and identification of implicit prejudice as consequential, ubiquitous, and distinct from "explicit," or conscious, endorsement of prejudiced attitudes has now been firmly established by decades of research, hundreds of studies, thousands of participants from around the world, and a variety of research methodologies. Implicit prejudice was captured initially in two basic experimental paradigms that emerged from the information-processing nexus of cognitive and social psychology—one demonstrating the effects of concepts made implicitly salient through experimental manipulation, and the other demonstrating the existence and correlates of implicit semantic associations.

The effects of cognitively salient concepts on social judgment were initially captured in now-classic experiments demonstrating that evaluations of social targets are implicitly influenced by recent exposure to judgment-related information (Higgins, Rholes, and Jones, 1977; Srull and Wyer, 1979). Although interdisciplinary consensus about the importance of implicit cognition exhibited by this research tradition had been building for many years, its application to stereotyping was captured in Patricia Devine's iconic paper (1989), which marked the beginning of a

paradigm shift in the social-psychological understanding of stereotyping and prejudice more generally.[1]

In the critical experiment, participants evaluated a hypothetical person named "Donald" as more hostile if they had been subliminally exposed to a large versus a small proportion of words related to common U.S. stereotypes of African Americans. The finding was striking because it suggested that crude stereotypes could operate unintentionally and outside conscious awareness to influence social judgment, and it was disturbing because it showed that implicit stereotyping occurred to an equal degree whether participants explicitly endorsed racist attitudes or not.

This basic paradigm has since been used in scores of experiments that confirm the implicit operation of prejudice and stereotyping in social judgment including, but not limited to, ethnicity and race (e.g., Dovidio et al., 1997), gender (e.g., Rudman and Borgida, 1995), and age (e.g., Levy, 1996). As an example of the existence of implicit gender stereotypes, women but not men were judged as more dependent after recent exposure to female stereotypes, and men but not women were judged as more aggressive after exposure to male stereotypes (Banaji, Hardin, and Rothman, 1993). The effects of stereotype salience were equally large for women and men, regardless of the levels of explicit prejudice. In sum, research in this tradition suggests that mere knowledge of a stereotype can influence social judgment regardless of explicit intentions and regardless of the social category of the one doing the stereotyping.

Research demonstrating the implicit influence of cognitively salient stereotypes in social judgment has been complemented by research in the second paradigm that establishes the extent to which stereotyping and prejudice operate as webs of cognitive associations. Like Freud's discovery that mental architecture is revealed by quantifying what most easily comes to mind given targeted conceptual probes, the notion was initially captured in now-classic experiments showing that judgments on "target" words are faster if they are immediately preceded by brief exposure to semantically related, as opposed to unrelated, "prime" words (e.g., Meyer and Schvaneveldt, 1971; Neely, 1976, 1977). These semantic relations are now known to be highly correlated with those identified in free-association tasks (for a review see Ratcliff and McKoon, 1994). Extensive research demonstrates that a variety of social beliefs and attitudes function as semantic and evaluative associations across several procedural variations, including conditions in which the prime words are exposed too quickly for people to see (for reviews see Fazio, 2001; Greenwald and Banaji, 1995). For example, simple judgments about target female pronouns were faster after brief exposure to prime words either denotatively or connotatively related to women (e.g., lady, nurse) than words related to men (e.g., gentleman, doctor), and judgments about male pronouns were faster after exposure to prime words related to men than women (Banaji and Hardin, 1996; Blair and Banaji, 1996). Similarly, people were faster to judge words associated with negative stereotypes of African Americans after exposure to black faces than to white faces (e.g., Dovidio, Evans, and Tyler, 1986; Dovidio et al., 1997; Wittenbrink, Judd, and Park, 1997). Such results have been taken to demonstrate the automatic nature of beliefs or stereotypes when they capture associations between social groups and their common stereotypes, and have been used to demonstrate the automatic nature of attitudes or preferences when they capture associations between social groups and common evaluations of them.

Research in this tradition suggests the ubiquity with which common prejudice and stereotyping operates among all kinds of people along lines laid down by extant social relations on a variety of dimensions. These include, but are not limited to, ethnicity and race (e.g., Nosek, Banaji, and Greenwald, 2002a), gender (e.g., Banaji and Hardin, 1996), sexual orientation (e.g., Dasgupta and Rivera, 2008), body shape (e.g., Bessenoff and Sherman, 2000), the elderly (Perdue and Gurtman, 1990), and adolescents (Gross and Hardin, 2007). Implicit prejudice of this kind develops early in children across cultures (e.g., Baron and Banaji, 2006; Dunham, Baron, and Banaji, 2006, 2007) and appears to involve specific brain structures associated with nonrational thought (e.g., Cunningham, Nezlek, and Banaji, 2004; Lieberman, 2000; Phelps et al., 2000).

Characteristics of Implicit Prejudice

Although the identification of the course, consequences, and nature of implicit prejudice continues to evolve in research spanning disciplines, research methodologies, and specific social categories, its fundamental characteristics are now firmly established. Implicit prejudice (a) operates unintentionally and outside awareness, (b) is empirically distinct from explicit prejudice, and (c) uniquely predicts consequential social judgment and behavior. Underlying all claims about the operation of implicit prejudice is the fact that the implicit operation of stereotypes and prejudice is robust and reliably measured, as indicated by hundreds of published experiments (e.g., Banaji, 2001; Greenwald and Banaji, 1995). In addition, research shows that implicit prejudice is subject to social influence, a finding that is important to public policy considerations, although the immediate operation of

implicit prejudice is difficult, if not impossible, to control through individual volition.

The most important characteristic of implicit prejudice is that it operates ubiquitously in the course of normal workaday information processing, often outside of individual awareness, in the absence of personal animus, and generally despite individual equanimity and deliberate attempts to avoid prejudice (for reviews see Devine, 2005; Dovidio and Gaertner, 2004). Evidence of this process includes experiments demonstrating that social judgment and behavior is affected in stereotype-consistent ways by unobtrusive, and even subliminal, manipulations of stereotype salience. Typically in these kinds of experiments, participants attempt to be fair and unbiased and, moreover, exhibit no evidence of knowing that their recent experience included exposure to stereotypes used in their evaluations. Experiments that manipulate stereotype salience subliminally through extremely rapid exposure to words or images make the case especially strongly (for reviews see Bargh, 1999; Devine and Monteith, 1999). Interestingly, implicit prejudice of this kind appears to operate regardless of the personal characteristics of research participants, including participant social category, and regardless of individual differences in related explicit attitudes and implicit attitudes. The implication is that anyone who is aware of a common stereotype is likely to use it when it is cognitively salient and relevant to the judgment at hand (e.g., Hardin and Rothman, 1997; Higgins, 1996).

Complementary evidence that prejudice operates implicitly comes from research using measures of automatic cognitive association, including serial semantic priming paradigms (e.g., Blair and Banaji, 1996), subliminal serial priming paradigms (e.g., Fazio et al., 1995), and startle responses (e.g., Amodio, Harmon-Jones, and Devine, 2003), as well as behavioral interference paradigms like Stroop tasks (e.g., Bargh and Pratto, 1986; Richeson and Trawalter, 2005) and implicit association tasks (IAT; e.g., Greenwald, McGhee, and Schwartz, 1998). Hundreds of experiments using these measures suggest that people are generally surprised to learn that they have implicit prejudices.

A second major characteristic of implicit prejudice is that it is difficult for individuals to deliberately modulate, control, or fake (for reviews see Devine and Monteith, 1999; Dovidio, Kawakami, and Gaertner, 2002; Greenwald et al., 2009). Experiments like Devine's (1989), which demonstrate implicit prejudice through subliminal, unconscious manipulations of stereotype salience, by design preclude individual awareness and control, thereby demonstrating that immediate conscious awareness of stereotyped information is formally unnecessary to produce implicit

stereotyping. Similar experiments that manipulate stereotype salience through recent conscious exposure to stereotyped information suggest that implicit stereotyping can occur through the kind of mere exposure to stereotyped information that occurs in the hurly-burly of everyday life in societies that are organized around race, class, and gender (e.g., Rudman and Borgida, 1995). Moreover, research expressly designed to test the success of individuals to control or fake their levels of implicit prejudice as assessed by measures of association show that it is extremely difficult or impossible to do so (Bielby, 2000), whether attitudes are about gays (e.g., Banse, Seise, and Zerbes, 2001), ethnic groups (e.g., Kim, 2003), or gender (e.g., Blair and Banaji, 1996).

Independent of individual attempts to control the operation of implicit prejudice, research shows that it is nearly impossible to consciously correct for effects of implicit prejudice (for one review see Wegener and Petty, 1997). To do so, one must be in the unlikely circumstance of having all at once (a) knowledge that implicit prejudice is operating, (b) both the motivation and cognitive capacity to control it, and perhaps most unlikely of all, (c) precise knowledge of the magnitude and direction of the correction needed (e.g., Bargh, 1999; Fazio and Towles-Schwen, 1999). For example, although individual differences in explicit prejudice predict the overt interpersonal friendliness of whites toward blacks, it is individual differences in implicit prejudice that predicts the nonverbal behavior of whites, which is the behavior that, in turn, predicts black attitudes toward whites (e.g., Dovidio, Kawakami, and Gaertner, 2002).

The third critical characteristic of implicit prejudice is that it is empirically distinct from explicit prejudice, including activating distinctive regions of the brain (Cunningham, Nezlek, and Banaji, 2004). Although explicit attitudes are often uncorrelated with the implicit operation of prejudice (e.g., Devine, 1989; Fazio and Olson, 2003) and implicit prejudiced associations (e.g., Gross and Hardin, 2007), correlations between implicit and explicit attitudes actually vary widely across studies (e.g., Hofmann et al., 2005; Nosek, 2005). A picture of when and why implicit and explicit attitudes are likely to be dissociated has begun to emerge. Baldly explicit prejudice on the basis of race and gender often conflicts with social norms of equity and justice and hence is a domain in which implicit-explicit attitude dissociations often occur. In contrast, in domains in which explicit attitudes do not conflict with consensual social norms, implicit and explicit attitudes are often correlated (e.g., Gawronski, 2002; Greenwald et al., 2009). For example, implicit prejudice is correlated with amygdala activation (Cunningham, Nezlek, and Banaji,

2004; Phelps et al., 2000), and explicit prejudice is more strongly correlated with prefrontal cortex activation (Cunningham et al., 2004; see also Amodio et al., 2004). Most importantly, implicit prejudice uniquely predicts related attitudes and behavior over and above explicit prejudice and appears to be related to distinct families of social judgment and behavior. Implicit attitudes are associated relatively more with tacit learning, manipulations, and consequences, whereas explicit attitudes are relatively more associated with intentionally controllable behaviors and attitudes (e.g., Olson and Fazio, 2003; Spalding and Hardin, 1999).

Because the unique predictive validity of implicit prejudice is critical to appreciating its implications for policy choices, we now turn to a detailed discussion of this evidence in the context of policy implications.

Consequences and Social Control of Implicit Prejudice

The existence of implicit prejudice would be of little practical consequence if it were an unreliable predictor of social judgment and behavior, particularly given the growing interest in its potential economic, labor, legal, and policy implications (e.g., Ayres, 2001; Banaji and Bhaskar, 2000; Banaji and Dasgupta, 1998; Chugh, 2004; Greenwald and Krieger, 2006; Jost et al., 2009; Kang and Banaji, 2006; Tetlock and Mitchell, in press). However, research demonstrates the consequential nature of implicit prejudice in a variety of domains, including health, job satisfaction, voting behavior, and social interaction. Our discussion of this evidence is organized around the two paradigms that led to the discovery of implicit prejudice in the first place—the implicit effects of cognitively salient stereotypes and prejudice, and the predictive utility of implicit associations between social groups and their presumed characteristics.

Implicit Effects of Cognitively Accessible Stereotypes and Prejudice

Perhaps the most disturbing aspect of implicit prejudice is that while cognitively salient stereotypes and prejudices operate outside of conscious awareness, they produce qualitative changes in social judgment and behavior. Across some two dozen experiments in which participants are presented with a series of images of social situations and instructed to as quickly and accurately as possible "shoot" if the target is armed and "don't shoot" if the target is unarmed, the finding is consistent: participants faster and more accurately

shoot gun-toting black targets than white targets and faster and more accurately avoid shooting tool-toting white targets than black targets (e.g., Correll et al., 2002; Correll, Urland, and Ito, 2006). The finding is obtained among both white and black participants alike, and even among professional police officers (Correll et al., 2007; Plant and Peruche, 2005; Plant, Peruche, and Butz, 2005). In a similar experimental paradigm in which participants were instructed to distinguish between weapons and hand tools, participants were faster to correctly identify weapons after exposure to black faces than to white faces but faster to correctly identify tools after exposure to white faces than to black faces (Payne, 2001). A follow-up study demonstrated that participants under time pressure were more likely to misidentify tools as guns after exposure to black faces but misidentify guns as tools after exposure to white faces (see also Govorun and Payne, 2006; Payne, Shimizu, and Jacoby, 2005), a finding that is obtained even among professional police officers (Eberhardt et al., 2004).

Such findings have important implications for police officers, given the broader finding that police consistently use greater lethal and nonlethal force against nonwhite suspects than white suspects (e.g., for reviews see U.S. Department of Justice, 2001; Geller, 1982). Indeed, Los Angeles police officers judge adolescents accused of shoplifting or assault more negatively and as more culpable when they have been subliminally exposed to words related to common stereotypes about blacks than words that are not related to the stereotypes (Graham and Lowery, 2004).

The implicit use of common stereotypes is not limited to issues of race but is also seen in matters of age and in instances of gender bias. For example, the behavior of a seventeen-year-old (but not a seventy-one-year-old) toward a police officer is judged as more rebellious after the latter's subliminal exposure to words related to common adolescent stereotypes than with exposure to words that are not, and the magnitude of the effect is unrelated to individual differences in explicit attitudes about adolescents (Gross and Hardin, 2007). And, in a telling experiment involving stereotypes commonly traded in mass media (e.g., beer ads featuring bikini-clad models), recent exposure to sexist versus nonsexist television advertisements was shown to cause men to (a) evaluate a job applicant as more incapable and unintelligent, (b) evaluate her as more sexually attractive and receptive, (c) make more sexual advances to her, and (d) evaluate her as more deserving of being hired (Rudman and Borgida, 1995). Here, too, typical of experiments of this type, the effect of exposure to sexist ads was unqualified by individual differences in explicit endorsement of sexist beliefs and attitudes.

Implicit prejudice and stereotyping is not limited to judgments of others, however, but also affects self-judgment and behavior, especially with regard to intellectual performance. For example, Asian American women believe they are relatively better at math than verbal skills when they have identified their ethnicity, but better at verbal than math skills when they have identified their gender (e.g., Sinclair, Hardin, and Lowery, 2006). Even more striking are findings that similar manipulations implicitly affect stereotype-related intellectual performance. Consistent with the respective stereotypes, blacks, but not whites, perform worse on GRE advanced exams when ethnicity is salient (e.g., Steele and Aronson, 1995), and women, but not men, perform worse on GRE quantitative exams (Spencer, Steele, and Quinn, 1999), and worse on a logic task but not an identical verbal task, when gender is salient (Cheung and Hardin, 2010). Similarly, older, but not younger people, perform worse on memory tasks when age is salient (e.g., Levy, 1996), and students from low, but not high, socioeconomic backgrounds perform worse on intellectual tasks when economic status is salient (e.g., Croizet and Claire, 1998; Harrison et al., 2006). Moreover, gender and ethnic stereotypes can interact to produce especially large decrements in the math and spatial performance of Latina women (e.g., Gonzales, Blanton, and Williams, 2002). Such performance discrepancies are also evident via functional magnetic resonance imaging (fMRI) data. For example, women not only perform worse on mental rotation tasks when negative stereotypes are salient but performance decrements are correlated with greater activity in brain regions associated with emotion and implicit prejudice (Wraga et al., 2007).

Congruent with evidence discussed throughout this paper, the consequences of implicit prejudice to the self echo the principled operation of implicit prejudice more generally. Stereotypes are double-edged swords and hence can sometimes boost performance. For example, Asian American women perform better on quantitative tests when their ethnicity is salient than when their gender is salient (e.g., Shih, Pittinsky, and Ambady, 1999). Whether positive or negative, implicit stereotype threat effects emerge early in development and appear with increasing strength throughout elementary and middle school (e.g., Ambady et al., 2001). Finally, evidence suggests that these kinds of effects are more likely to occur when the relevant stereotypes are made salient in subtle ways rather than blatantly (Shih et al., 2002), congruent with our broader argument about the insidious role that implicit prejudice plays in everyday social cognition and behavior.

Implicit Prejudice as Cognitive Associations

Common stereotypes and prejudice not only affect social judgment and behavior implicitly, but several measures of implicit attitudes have been developed (for reviews see Olson and Fazio, 2003; Wittenbrink and Schwartz, 2007), and research based on hundreds of studies shows that implicit attitude measures are stable over time, internally consistent, and reliably predict related judgments and behaviors, including political attitudes, voting, academic achievement scores, consumer preferences, social evaluation, hiring decisions, and verbal and nonverbal affiliation (for reviews see Fazio and Olson, 2003; Nosek, 1995; Perugini, 2005). According to a recent meta-analysis (Greenwald et al., 2009), although implicit and explicit attitudes are commonly uncorrelated with each other, implicit measures are, on average, comparably correlated with criterion measures and usually more strongly correlated with measures of socially sensitive behavior than explicit measures. In short, where stereotyping and prejudice are concerned, implicit measures generally predict behavior better than explicit measures.

Unlike explicit measures, in which predictive validity often declines substantially for socially sensitive criteria, the predictive validity of implicit measures typically does not. For example, in a study reported by Rudman and Ashmore (2007), implicit prejudice uniquely predicts self-reported hostile behavior among whites toward blacks, including ethnic slurs, ostracism, and verbal and physical abuse, and does so over and above explicit attitudes and prejudice. In a second study, implicit prejudice among whites toward Jews, Asians, and blacks was shown to predict preferences to de-fund campus organizations representing Jews, Asians, and blacks, respectively—again, over and above explicit attitudes and prejudice. Implicit prejudice can also predict prejudice-related judgments when explicit attitudes do not, particularly in cases of intergroup relations (reviewed in Greenwald et al., 2009). For example, unlike explicit prejudice, implicit racial prejudice among whites predicts quickness to perceive anger in black faces but not white faces (Hugenberg and Bodenhausen, 2003).

It is one thing for individual differences in implicit prejudice to predict attitudes and judgment, but it is quite another for it to predict behavior. Implicit attitudes predict nonverbal friendliness and discomfort of whites when interacting with blacks (Dovidio et al., 1997, 2002) and how positively blacks perceive whites with whom they interact (Dovidio, Kawakami, and Gaertner, 2002; Fazio et al., 1995; Sekaquaptewa et al., 2003). For example, in research particularly

telling for common educational and school situations, Richeson and Shelton (2005) found that in face-to-face interpersonal interactions, individual differences in implicit prejudice were more apparent to black than white perceivers and more apparent when whites interacted with blacks than with other whites (see also Perugini, O'Gorman, and Prestwich, 2007; Ziegert and Hanges, 2005).

Implicit attitudes not only affect social judgment and behavior relative to others but also are important predictors of one's own behavior and self-evaluation. For example, implicit, but not explicit, self-esteem predicts anxious behavior in self-threatening situations but not in unthreatening situations (Spalding and Hardin, 1999; see also Asendorpf, Banse, and Mucke, 2002; Egloff and Schmukle, 2002). Women who implicitly associate romance with chivalry report less interest in economic and educational achievement (Rudman and Heppen, 2003), and implicit dissociations between the concepts of math and women predict lower quantitative SAT scores among women (Nosek, Banaji, and Greenwald, 2002b). Finally, a surprising number of African Americans exhibit implicit preference for whites over blacks (e.g., Nosek, Banaji, and Greenwald, 2002a). Variability in implicit antiblack prejudice among African Americans predicts stated preferences for working with white versus black partners on intellectually demanding tasks and does so independently of explicit attitudes (Ashburn-Nardo, Knowles, and Monteith, 2003), a finding suggesting that the general tendency to favor in-groups over out-groups may be trumped by implicit stereotypes relevant to the task at hand (see also Rudman, Feinberg, and Fairchild, 2002).

Most of the research on the predictive validity of implicit prejudice discussed thus far involves undergraduate participant samples in laboratory settings, yet one might rightly wonder whether implicit prejudice will matter in daily tasks, big and small. One reason to believe that it will is research showing that among people who have finished their formal education, implicit attitudes predict behavior and judgment on dimensions that matter to people beside college students and do so on a variety of dimensions of undeniable real-world application. For example, implicit attitudes predict suicide attempts (Glashouwer et al., 2010; Nock and Banaji, 2007; Nock et al., 2010), severity and treatment outcomes for phobia and panic disorders (e.g., Teachman, Marker, and Smith-Janik, 2008; Teachman, Smith-Janik, and Saporito, 2007; Teachman and Woody, 2003), condom use (Marsh, Johnson, and Scott-Sheldon, 2001), smoking status (Swanson, Rudman, and Greenwald, 2001), alcohol consumption (Weirs et al., 2002), and consumer preferences for consumer goods like yogurt, beverages, and fast-food restaurants (Maison, Greenwald, and Bruin, 2004). In addition, reductions in implicit romantic attraction predict the subsequent breakup of committed relationships (Lee, Rogge, and Reis, 2010).

In addition to the large and growing literature demonstrating the predictive validity of measures of implicit attitudes in matters of everyday life, research shows that implicit prejudice predicts behavior outside the laboratory. For example, implicit preference among Swedish job recruiters for native Swedes over Arabs predicts interview preferences (Rooth, 2010). Overall, native Swedes were more than three times more likely to receive interview callbacks than equally qualified Arabs.

Several studies demonstrate that implicit prejudice predicts voting behavior, including the historic 2008 election in which Barack Obama became the first African American to be elected president of the United States. For example, in the week before the election, implicit antiblack prejudice predicted intention to vote for John McCain over Obama and did so independently of self-reported conservatism (Greenwald et al., 2009). Another study found that the degree to which participants implicitly associated America more with McCain than Obama predicted intention to vote for McCain (Devos and Ma, 2010).

Implicit prejudice not only predicts voting intentions before elections but also reported voting behavior after elections. Voters were substantially less likely to report voting for Barack Obama, and exhibited more negative attitudes toward health care reform, the greater their implicit prejudice (Knowles, Lowery, and Shaumberg, 2010), and, in a follow-up study conducted nearly a year after the election, implicit prejudice remained a significant predictor of negative attitudes toward Obama. Moreover, implicit prejudice predicted negative attitudes about health-care reform when it was ascribed to Obama but not when the identical reform was ascribed to Bill Clinton. Similar findings have obtained in studies of the Italian electorate, as well (e.g., Arcuri et al., 2008; Galdi, Arcuri, and Gawronski, 2008; Roccato and Zogmaister, 2010).

Another area of society in which the real-world operation of implicit prejudice is implicated is in the practice of medicine, in which differential treatment as a function of ethnicity is a well-documented case in point. A recent study of emergency-room treatment of more than 150,000 patients complaining of severe pain over a 13-year span found that whites were given powerful opioid pain killers more than blacks and Hispanics, with evidence suggesting that

the disparity is due more to undertreatment of minorities rather than overtreatment of whites (Pletcher et al., 2008). Racial disparities are well documented for treatment of cardiovascular disease as well (for a review see Kressin and Petersen, 2001), including expensive treatments for acute myocardial infarction (e.g., Petersen et al., 2002).

New evidence suggests that at least one cause for such findings may be individual differences in implicit prejudice among treating physicians. In a study that assessed both explicit and implicit attitudes toward whites and blacks and treatment recommendations for hypothetical patients who differed only as a function of an experimental manipulation of race, emergency-room physicians exhibited strong implicit preference for whites over blacks, and also strong implicit associations of blacks versus whites for being uncooperative, despite exhibiting no explicit preferences for whites or differences in cooperativeness between whites and blacks. Importantly, however, although explicit attitudes did not predict emergency treatment recommendations, implicit attitudes did. Greater implicit prejudice predicted an increasing likelihood to treat whites and a decreasing likelihood to treat blacks exhibiting identical symptoms (Green et al., 2007). By extension, and perhaps unsurprisingly, implicit racial bias among physicians negatively predicts African American patient satisfaction with their physicians (Penner et al., 2010).

Consistent with laboratory findings suggesting that implicit attitudes should be uniquely strong predictors of counternormative behavior, implicit negative attitudes toward injection-drug users among drug and alcohol nurses who treat them predicts nurses' stated intentions to leave drug and alcohol nursing, over and above relevant explicit attitudes (von Hippel, Brener, and von Hippel, 2008),[2] corroborating laboratory demonstrations of the unique predictive power of implicit measures when judgments are potentially nonnormative (Greenwald et al., 2009). In other words, although the medical model frames drug and alcohol abuse as an involuntary disease to be treated, and as such abusers should be worthy of sympathy, the day-to-day experience with a population known to be difficult and challenging by a part of the medical community that is known to have a high job turnover rate may make expressly negative attitudes about abusers counternormative. In addition, it is implicit prejudice (but not explicit prejudice) that mediates the well-documented relation between stress and intention to change jobs (von Hippel, Brener, and von Hippel, 2008).

In short, research demonstrating the real-world applicability of implicit attitudes continues to grow, and it is no longer credible to hide behind the view that the predictive validity of implicit prejudice on judgment and behavior is a quirk of the laboratory (see also Jost et al., 2009).

Social Control of Implicit Prejudice

Given evidence that implicit prejudice is reliably captured and measured and that it is consequential, ubiquitous, and stubbornly immune to individual attempts to control it, what hope is there for effective policy solutions? Although implicit prejudice presents challenges to public policy formulations based on outdated notions of the nature of prejudice, recent research shows that it behaves in predictable ways that conform to fundamental principles of social and cognitive psychology. Implicit prejudice reflects stable social relationships and organization by reflecting social identities, group categorizations and status, as well as general preferences for the self, similar others, and in-groups (e.g., Bosson, Swann, and Pennebaker, 2000; Greenwald, McGhee, and Schwartz, 1998; Spalding and Hardin, 1999). Moreover, evidence suggests that implicit prejudice is responsive to social dynamics, including (a) relative intergroup status (e.g., Rudman, Feinberg, and Fairchild, 2002), (b) minimal group categorization (Ashburn-Nardo, Voils, and Monteith, 2001), (c) chronic and temporary changes in the salience of prejudice-related information (e.g., Dasgupta and Greenwald, 2001), and (d) friendly intergroup contact (e.g., Tam et al., 2006). Implicit prejudice can also increase and decrease as a function of conditioning that is consistent with the fundamentals of learning theory (e.g., Bargh, 1996; Fazio 2001, 2003; Fazio and Olson, 2003; Hardin and Rothman, 1997), and it generally conforms to principles of cognitive consistency (e.g., Greenwald et al., 2009).

An obvious but important indication of the way implicit prejudice reflects social dynamics is the fact that it so well tracks the character of chronic social organization, including relative group power, social status, and concomitant stereotypes. For example, although in-group preference is a common feature of implicit prejudice (e.g., Greenwald et al., 1998), at least as important are findings that it reflects social status. Members of high-status groups in the United States not only exhibit greater implicit group favoritism than low-status groups but also do so as a function of their relative status, whether they are rich, white, skinny, or Christian (e.g., Nosek et al., 2002a; Rudman, Feinberg, and Fairchild, 2002). However, at the same time, although in-group preference is common in both implicit and explicit prejudice, out-group preference is hardly rare (e.g., Jost and Banaji, 1994) and also closely aligns with relative group

status. For example, members of low-status groups were more likely to implicitly favor dominant out-groups to the extent that their in-group was low in status, despite exhibiting strong explicit in-group favoritism (Jost, Pelham, and Carvallo, 2002; Rudman, Feinberg, and Fairchild, 2002).

Implicit prejudice not only reflects stable social and organizational hierarchies, but research shows that changes in social organization also predict corresponding changes in implicit prejudice, a finding that has promising implications for public policy. Friendly intergroup contact is shown to reduce both implicit and explicit prejudice alike (e.g., Henry and Hardin, 2006; Turner, Hewstone, and Voci, 2007). In one example, implicit prejudice toward gay and lesbian people was found to be lower for people who reported high levels of long-term contact with gay and lesbian people as well as for people who reported being exposed to gay-positive media (Cheung et al., 2011; Dasgupta and Rivera, 2008). Similarly, implicit prejudice toward the elderly was lower among college students the more friendships they reported having with older people (Tam et al., 2006). In yet another example, implicit prejudice was found to be lower between British and South Asian children in England to the extent that they reported out-group friendships, and implicit prejudice was reduced even among children who reported no out-group friendships themselves but who reported having friends who did (Turner, Hewstone, and Voci, 2007). Causal modeling in this research indicates that the findings are more consistent with intergroup friendships affecting implicit prejudice than with implicit prejudice affecting friendship patterns (Tam et al., 2006; Turner, Hewstone, and Voci, 2007), a conclusion corroborated experimentally. For example, implicit prejudice among white college freshmen was reduced more over the course of their first school term if they were randomly assigned to a black roommate than a white roommate (Shook and Fazio, 2007).

Although friendly intergroup contact generally reduces implicit intergroup prejudice, recent findings demonstrate that intergroup contact does not always have purely positive outcomes. For example, anti-adolescent implicit prejudice among adolescents was greater to the degree that they reported having close friendships with adults (Gross and Hardin, 2007). Evidence also suggests that relatively stable aspects of social hierarchy complicate matters. In research involving blacks and whites in Chicago and Christians and Muslims in Lebanon, implicit intergroup prejudice was shown to be lower to the degree that participants reported out-group friendships (Henry and Hardin, 2006). However, results also indicate that implicit prejudice reduction is greater for low-status group members toward high-status group members than it is for high-status group members toward low-status group members. That is, in this study, out-group friendships predicted greater reductions in implicit prejudice for Muslims than Christians and for blacks than whites due to their places in the social hierarchy.

Research also indicates that implicit prejudice is affected by social dynamics throughout development (e.g., Baron and Banaji, 2006; Rutland et al., 2005) and that the development of implicit prejudice is likely to be bound up with interpersonal dynamics involving interpersonal identification and inter-subjectivity (e.g., Hardin and Conley, 2001; Hardin and Higgins, 1996). For example, implicit intergroup prejudice between Korean and Japanese students in the United States was greater to the degree that participants remained connected to their ethnic heritage as indicated by linguistic fluency (Greenwald, McGhee, and Schwartz, 1998). People exhibited more positive implicit attitudes toward women to the degree that they reported being raised more by their mothers than their fathers (Rudman and Goodwin, 2004). And, implicit racial prejudice among white fourth- and fifth-grade children was correlated with the explicit prejudice of their parents, but only to the extent that they identified with their parents (Sinclair, Lowery, and Dunn, 2005), and the implicit prejudice of mothers predicted racial preferences exhibited by their three- to six-year-old children (Castelli, Zogmaister, and Tomelleri, 2009).

Research demonstrating the long-term social determinants of implicit prejudice is likely to be either encouraging or depressing, depending upon one's sense of the likelihood of broad, long-term changes in social organization and culture. It is important, however, to remember that such things do happen. What changes in implicit prejudice might be revealed if the measures had been in existence long enough to reflect suffrage, women's mass entry into the workforce during World War II, the civil rights movement, and twentieth-century urban white flight, to name just a few societal sea changes?

Although we believe that culture-wide changes in implicit prejudice will require culture-wide changes in social organization and practice, another way in which implicit prejudice obeys principles of social psychology offers some promise of more immediate, if local, opportunities for progress. Research shows that implicit prejudice is subject to the demands of immediate situations and interpersonal dynamics, much like human behavior more generally (e.g., Ross and Nisbett, 1991). For example, white participants exhibited lower implicit prejudice in the presence of a black experimenter than a white experimenter

(Lowery, Hardin, and Sinclair, 2001; Richeson and Ambady, 2003). Interestingly, however, Lowery and colleagues (2001) also found that this automatic social tuning effect did not occur among Asian American participants, whose implicit prejudice was reduced only when the experimenter expressly told them to avoid prejudice. This finding suggests that although the norm to avoid prejudice may operate tacitly for some, it may require explication for people who do not yet recognize their potential role as ciphers of prejudice.

Research also suggests that the interpersonal regulation of implicit prejudice is due in part to a motivation to affiliate with others who are presumed to hold specific values related to prejudice, as implied by shared reality theory (e.g., Hardin and Conley, 2001). For example, participants exhibited less implicit racial prejudice in the presence of an experimenter wearing a T-shirt with an antiracism message than a blank T-shirt, but only when the experimenter was likeable (Sinclair et al., 2005). When the experimenter was not likeable, implicit prejudice was actually greater in the presence of the ostensibly egalitarian experimenter. In addition, social tuning in these experiments was mediated by the degree to which participants liked the experimenter, providing converging evidence that interpersonal dynamics play a role in the modulation of implicit prejudice, as they do in other dimensions of social cognition (Hardin and Conley, 2001; Hardin and Higgins, 1996).

As regards public and personal policy, these findings suggest that a public stance for egalitarian values is a double-edged sword, and a sharp one at that. Although it may reduce implicit prejudice among others when espoused by someone who is likeable and high in status, it may backfire when espoused by someone who is not likeable or otherwise of marginal status. This finding suggests one mechanism by which common forms of "sensitivity training" in service of the reduction of workplace sexism and racism may be subverted by interpersonal dynamics, however laudable the goals.

Demonstrating the utility of specific interventions to reduce implicit prejudice, Rudman, Ashmore, and Gary (2001) found that diversity education with a likeable black professor reduced implicit prejudice and did so through liking for the professor, increased friendships with other African Americans, and reduced fear of blacks. Likewise, thinking about gay-positive role models reduced implicit prejudice for those with low contact with gay and lesbian people to the level of those with high contact and increased the endorsement of gay-positive attitudes, including legalizing civil unions for gays and lesbians (Dasgupta and Rivera, 2008).

In a cautionary note, however, the lack of long-term exposure to a particular group can sometimes trigger greater implicit prejudice when a member of the group is present. In one example, people who reported having no gay friends at all exhibited greater implicit antigay prejudice when a male experimenter incidentally mentioned his "boyfriend" than when he mentioned his "girlfriend." Similarly, women who reported having no lesbian friends exhibited greater implicit antilesbian bias when the experimenter was from a gay and lesbian organization (Cheung et al., 2011). This research complements research showing immediate social influence on implicit prejudice. It suggests that as powerful as immediate social norms might be, implicit prejudice is ultimately expressed differently from individual to individual as a function of attitudes presumed to be held by others in relevant long-term social relationships, sometimes in subtle or even contradictory ways, much as it depends on other dimensions of social cognition (e.g., Hardin and Higgins, 1996).

Research demonstrating that implicit prejudice is subject to social influence is broadly consistent with principles of information processing (for a review see Blair, 2002). Implicit racial prejudice is reduced (a) when admired black exemplars are used (e.g., Dasgupta and Greenwald, 2001; cf. De Houwer, 2001), (b) after seeing an image of blacks at a friendly barbeque versus unfriendly street corner (Wittenbrink, Judd, and Park, 2001), and (c) imagining the virtues of multicultural education (Richeson and Nussbaum, 2004). In contrast, implicit racial prejudice is increased after exposure to violent rap music (Rudman and Lee, 2002). Implicit gender stereotyping is reduced for those who have recently been exposed to images of female leaders (Dasgupta and Asgari, 2004) or have recently imagined a powerful woman (Blair, Ma, and Lenton, 2001). This research suggests that simple images and text in immediate situations can affect levels of implicit prejudice for those in the situation in ways that are broadly congruent with construct accessibility theory (e.g., Bargh, 1996), which is the "common language" that underlies most information-processing theory in social cognition (Higgins, 1996).

Taken together, research on the social control of implicit prejudice is broadly congruent with the Marxian maxim that egalitarian societies elicit egalitarian-minded people, as well as with the Skinnerian maxim that admirable individual behavior is elicited by situations that reinforce admirable behavior. Indeed, the methodological and theoretical advances that have transformed the understanding of the nature of prejudice—including sometimes-puzzling relations between implicit and explicit

prejudice—resonates with what Skinner argued about the relation between scientific advances and the understanding of human nature more generally:

> The line between public and private is not fixed. The boundary shifts with every discovery of a technique for making private events public . . . The problem of privacy may, therefore, eventually be solved by technical advance.
>
> —*B.F. Skinner, 1953, p.282*

Conclusions

It is not far-fetched to argue that successful policy solutions to the problem of prejudice are best pursued in light of the science of the nature of prejudice. Research in recent decades has revealed the insidious capacity of prejudice to operate implicitly—unwittingly, unintentionally, and unavoidably—as well as its course, consequences, and control at the nexus of individual cognition and social relations. In some ways, the transformative understanding of the nature of prejudice brings full circle the story of human nature since its inception in American social psychology in the mid-twentieth-century work of Sherif, Lewin, Asch, and others as an attempt to understand how seemingly good people can participate in genocide, which is also captured in Hannah Arendt's memorable phrase, "the banality of evil."

Indeed, the most important thing to know about the nature of prejudice is that it is ever present in human behavior and cognition. It remains sufficiently in the background such that it eludes conscious awareness and immediate individual control, yet it is often consequential in everyday life. Its capacity to affect social judgment and behavior without personal animus or hostility is dismissed or ignored at some peril, because a continued focus on the problem of prejudice as a result of the nonnormatively hostile behavior of the few is likely to distract policy makers from adopting strategies more strongly rooted in the science of the many. What remains are questions about how best to deal with these discoveries in shaping personal and public policy—questions that are in this light only beginning to receive the empirical attention they deserve.

What must enter into any policy computation are additional facts about the nature of prejudice beyond the primary idea that banality is its modus operandi. We must add to this the idea that prejudices and stereotypes are rooted in social consensus; they are not random. Within a given society, the likes, dislikes, and beliefs that constrain some and privilege others occur in patterns that systematically oppress subordinates

while further ingraining the superiority of the dominants. Were the effects of prejudice and stereotypes less systematic, policy intervention would be less needed because their effects may be said to cancel each other out. However, when, for example, over 80% of American whites and Asians show antiblack bias and over 90% of Americans show anti-elder bias, we must pay heed. Policies that are willing to take into account the presence of implicit forms of prejudice and discrimination as a given will be the more forward-thinking instruments for change because they will be rooted in a truth about human nature and social contexts.

Furthermore, for societies that derive their sense of good character on the basis of personal accomplishment and meritocracy, research on implicit prejudice poses particularly thorny problems. The research we reviewed suggests that behavior is shaped by the social jostling and "sloshing around" of the individual, unbeknownst to the person and those around her, suggesting that the problem of implicit prejudice may be especially insidious in a society that celebrates, evaluates, and is organized around individual meritocracy. Indeed, research shows that beliefs in meritocracy pose special problems for members of stigmatized groups (e.g., Jost and Burgess, 2000; Jost and Thompson, 2000). For example, Filipina domestic workers in Hong Kong, as well as women in the United States, devalued the monetary value of their work more if their group identity was salient, but do so only to the degree that they endorsed system-justifying attitudes related to meritocracy (Cheung and Hardin, 2010). The aggregation of these kinds of effects, both large and small, but systematically organized across situations and social roles, suggests at the very least the possibility that even incrementally small biases may be expressed through actions that create a large divide among people.

Research demonstrating the effects of stereotypes and prejudice on behavior give direction to policy makers for the types of behavior most in need of their attention. It is our contention that locating the problem of prejudice in a few problematic individuals and designing solutions to the problem around this view is to miss the point. The profound implication of the discovery of implicit prejudice is that anybody is capable of prejudice, whether they know it or not, and of stereotyping, whether they want to or not. Therefore, given the implicit operation of prejudice and stereotyping and its ubiquitous nature, we believe that solutions should focus on identifying the enabling conditions that call out prejudice and stereotyping across individuals rather than focusing on identifying the rotten apples. Once identified, we must focus on the enabling conditions that promote egalitarianism

and healthy individuation. What kinds of situations bring out implicit egalitarian attitudes? Congruent with well-documented principles identified across the behavioral and mind sciences and corroborated in research on implicit prejudice, social situations populated with powerful, likeable people who are known or assumed to hold egalitarian values implicitly call out like minds in those around them.

Notes

1. Here and throughout we adopt conventions of social-psychological nomenclature in our use of terms. The umbrella term *attitude* includes evaluations (prejudice), beliefs (stereotypes), and behaviors (discrimination) regarding an attitude object. The terms *explicit* and *implicit* are used to capture a well-accepted heuristic dichotomy between modes of mental functions that operate largely consciously and reflectively versus unconsciously and automatically. Hence, *implicit attitude* refers to the strength of automatic association between an attitude object and characteristic attributes, *implicit prejudice* refers to the strength of automatic associations between social groups and attributes good and bad, and *implicit stereotyping* refers to the strength of automatic associations between social groups and characteristic attributes which may vary in evaluative valence.

2. Specific intention to change jobs is the strongest known predictor of actual voluntary job changes (van Breukelen, van der List, and Steensma, 2004).

References

Allport, G. W. (1958). *The nature of prejudice*. New York: Doubleday.

Ambady, N., Shih, M., Kim, A., and Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, *12*, 385–390.

Amodio, D. M., Harmon-Jones, E., and Devine, P. G. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *Journal of Personality and Social Psychology*, *84*, 738–753.

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., and Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science*, *15*, 88–93.

Arcuri, L., Castelli, L., Galdi, S., Zozmaister, C., and Amadori, A. (2008). Predicting the vote: Implicit attitudes as predictors of the future behavior of the decided and undecided voters. *Political Psychology*, *29*, 369–387.

Asendorpf, J. B., Banse, R., and Mucke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology*, *83*, 380–393.

Ashburn-Nardo, L., Knowles, M. L., and Monteith, M. J. (2003). Black Americans' implicit racial associations and their implications for intergroup judgment. *Social Cognition*, *21*, 61–87.

Ashburn-Nardo, L., Voils, C. I., and Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology*, *81*, 789–799.

Ayres, I. (2001). *Pervasive prejudice? Unconventional evidence of race and gender discrimination*. Chicago: University of Chicago Press.

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. I. Roediger and J. S. Nairne (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.

Banaji, M. R., Bazerman, M., and Chugh, D. (2003, December). How (un)ethical are you? *Harvard Business Review*, pp. 56–64.

Banaji, M. R., and Bhaskar, R. (2000). Implicit stereotypes and memory: The bounded rationality of social beliefs. In D. L. Schacter and E. Scarry (Eds.), *Memory, brain, and belief* (pp. 139–175). Cambridge, MA: Harvard University Press.

Banaji, M. R., and Dasgupta, N. (1998). The consciousness of social beliefs: A program of research on stereotyping and prejudice. In V. Y. Yzerbyt, G. Lories, and B. Dardenne (Eds.), *Metacognition: Cognitive and social dimensions* (pp. 157–170). Thousand Oaks, CA: Sage.

Banaji, M. R., and Greenwald, A. G. (1994). Implicit stereotyping and prejudice. In M. P. Zanna and J. M. Olson (Eds.), *The psychology of prejudice: The Ontario Symposium* (Vol. 7, pp. 55–76). Hillsdale, NJ: Lawrence Erlbaum.

Banaji, M. R., and Hardin, C. D. (1996). Automatic gender stereotyping. *Psychological Science*, *7*, 136–141.

Banaji, M. R., Hardin, C., and Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, *65*, 272–281.

Banse, R., Seise, J., and Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift fur Experimentelle Psychologie*, *48*, 145–160.

Bargh, J. A. (1996). Principles of automaticity. In E. T. Higgins and A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 169–183). New York: Guilford.

———. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken and Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). New York: Guilford.

Bargh, J. A., and Pratto, F. (1986). Individual construct

accessibility and perceptual selection. *Journal of Experimental Social Psychology*, *49*, 1129–1146.

Baron, A. S., and Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6, 10 and adulthood. *Psychological Science*, *17*, 53–58.

Bessenoff, G. R., and Sherman, J. W. (2000). Automatic and controlled components of prejudice toward fat people: Evaluation versus stereotype activation. *Social Cognition*, *18*, 329–353.

Bielby, W. T. (2000). Minimizing workplace gender and racial bias. *Contemporary Sociology*, *29*, 120–129.

Blair, I. V. (2002). The malleability of automatic stereotyping and prejudice. *Journal of Personality and Social Psychology Review*, *6*, 242–261.

Blair, I. V., and Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, *70*, 1142–1163.

Blair, I. V., Ma, J. E., and Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*(5), 828–841. doi:10.1037/0022-3514.81.5.828

Bobo, L. (2001). Racial attitudes and relations at the close of the twentieth century. In N. Smelser, W. J. Wilson, and F. Mitchell (Eds.), *America becoming: Racial trends and their consequences* (pp. 262–299). Washington, DC: National Academy Press.

Bosson, J. K., Swann, W. and Pennebaker, J. W. (2000). Stalking the perfect measure of self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, *79*, 631–643.

Castelli, L., Zogmaister, C., and Tomelleri, S. (2009). The transmission of racial attitudes within the family. *Developmental Psychology*, *45*, 586–591.

Cheung, R. M., Fingerhut, A., Johnson, A., Noel, S., Drus, M., and Hardin, C. D. (2011). *Religiosity, heterosexuality, and anti-gay prejudice: Shared norms in everyday social tuning*. Unpublished manuscript, Department of Psychology, Brooklyn College, CUNY.

Cheung, R. M., and Hardin, C. D. (2010). Costs and benefits of political ideology: The case of economic self-stereotyping and stereotype threat. *Journal of Experimental Social Psychology*, *46*(5), 761–766. doi:10.1016/j.jesp.2010.03.012

Chugh, D. (2004). Societal and managerial implications of implicit social cognition: Why milliseconds matter. *Social Justice Research*, *17*, 203–222.

Correll, J., Park, B., Judd, C. M., and Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*, 1314–1329.

Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., and Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology*, *92*, 1006–1023.

Correll, J., Urland, G. L., and Ito, T. A. (2006). Event-related potentials and the decision to shoot: The role of threat perception and cognitive control. *Journal of Experimental Social Psychology*, *42*, 120–128.

Croizet, J., and Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, *24*, 588–594.

Crosby, F., Bromley, S., and Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review. *Psychological Bulletin*, *87*, 546–563.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., and Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science*, *15*(12), 806–813. doi:10.1111/j.0956-7976.2004.00760.x

Cunningham, W. A., Nezlek, J. B., and Banaji, M. R. (2004). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. *Personality and Social Psychology Bulletin*, *30*(10), 1332 –1346. doi:10.1177/0146167204264654

Darley, J. M., and Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, *44*, 20–23.

Dasgupta, N., and Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on automatic gender stereotyping. *Journal of Experimental Social Psychology*, *40*, 642–658.

Dasgupta, N., and Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814.

Dasgupta, N., and Rivera, L. M. (2008). When social context matters: The influence of long-term contact and short-term exposure to admired outgroup members on implicit attitudes and behavioral intentions. *Social Cognition*, *26*, 54–66.

De Houwer, J. (2001). A structure and process analysis of the IAT. *Journal of Experimental Social Psychology*, *37*, 443–451.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.

———. (2005). Prejudice with and without compunction: Allport's inner conflict revisited. In J. F. Dovidio, P. Glick, and L. A. Rudman (Eds.), *On the nature of prejudice: Fifty years after Allport* (pp. 327– 342). Oxford: Blackwell.

Devine, P. G., and Monteith, M. J. (1999). Automaticity and control in stereotyping. In S. Chaiken and Y. Trope (Eds.), *Dual-process models and themes in social and cognitive psychology* (pp. 339–360). New York: Guilford Press.

Devos, T., and Ma, D. S. (2010). *How "American" is Barack Obama? The role of national identity in a historic bid for the White House.* Unpublished manuscript, Department of Psychology, University of Chicago.

Dovidio, J. F. (2001). On the nature of contemporary prejudice: The third wave, *Journal of Social Issues*, *57*, 829–849.

Dovidio, J. F., Evans, N. and Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, *22*, 22–37.

Dovidio, J. F., and Gaertner, S. L. (2004). Aversive racism. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 36, pp. 1–51). San Diego, CA: Academic Press.

Dovidio, J. F., Kawakami, K., and Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*, 62–68.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., and Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*, 510–540.

Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology*, *34*, 590–598.

Dunham, Y., Baron, A. S., and Banaji, M. R. (2006). From American city to Japanese village: A cross-cultural investigation of implicit race attitudes. *Child Development*, *77*, 1268–1281.

———. (2007). Children and social groups: A developmental analysis of implicit consistency among Hispanic-Americans. *Self and Identity*, *6*, 238–255.

Duraisingam, V., Pidd, K., Roche, A. M., and O'Connor, J. (2006). *Stress, satisfaction and retention among alcohol and other drug workers in Australia.* Adelaide, Australia: National Centre for Education and Training on Addiction.

Eberhardt, J. L., Goff, P. A., Purdie, V. J., and Davies, P. G. (2004). Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology*, *87*, 876–893.

Egloff, B., and Schmukle, S. C. (2002). Predictive validity of an Implicit Association Test for assessing anxiety. *Journal of Personality and Social Psychology*, *83*, 1441–1455.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, 15, 115–141.

———. (2003). Variability in the likelihood of automatic attitude activation: Data reanalysis and commentary on Bargh, Chaiken, Govender and Pratto (1992). *Journal of Personality and Social Psychology*, *64*, 753–758.

Fazio, R. H., Jackson, J. R., Dunton, B. C., and Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–1027.

Fazio, R. H., and Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology*, *54*, 297–327.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., and Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229–238.

Fazio, R. H., and Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken and Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 97–116). New York: Guilford.

French, R. M., and Cleeremans, A. (2002). *Implicit learning and consciousness: An empirical, philosophical, and computational consensus in the making.* Hove, UK: Psychology Press.

Galdi, S., Arcuri, L., and Gawronski, B. (2008). Automatic mental associations predict future choices of undecided decision-makers. *Science*, *321*, 1100–1102.

Gallon, S. L., Gabriel, R. M., and Knudsen, J.R.W. (2003). The toughest job you'll ever love: A Pacific Northwest treatment workforce survey. *Journal of Substance Abuse Treatment*, *24*, 183–196.

Gawronski, B. (2002). What does the implicit association test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology*, *49*, 171–180.

Geller, W. A. (1982). Deadly force: What we know. *Journal of Police Science and Administration*, *10*, 151–177.

Glashouwer, K. A., de Jong, P. J., Penninx, B.W.J.H., Kerkhof, A.J.F.M., van Dyck, R., and Ormel, J. (2010). Do automatic self-associations relate to suicidal ideation? *Journal of Psychopathology and Behavioral Assessment*, *32*, 428–437. doi:10.1007/s10862-009-9156-y

Gonzales, P. M., Blanton, H., and Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin*, *28*, 659–670.

Govan, C. L., and Williams, K. D. (2004). Changing the affective valence of stimulus items influences the IAT by re-defining the category labels. *Journal of Personality and Social Psychology*, *40*, 357–365.

Govorun, O., and Payne, B. K. (2006). Ego depletion and prejudice: Separating automatic and controlled components. *Social Cognition*, *24*, 111–136.

Graham, S., and Lowery, B. S. (2004). Priming unconscious racial stereotypes about adolescent offenders. *Law and Human Behavior*, *28*, 483–504.

Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., and Banaji, M. R. (2007).

Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, *22*, 1231–1238.

Greenwald, A. G., and Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.

Greenwald, A. G., and Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review*, *94*, 945–967.

Greenwald, A. G., McGhee, D. E., and Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., and Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41.

Greenwald, A. G., Smith, C. T., Sriram, N., Bar-Anan, Y., and Nosek, B. A. (2009). Implicit race attitudes predicted vote in the 2008 U.S. presidential election. *Analyses of Social Issues and Public Policy*, *9*(1), 241–253. doi:10.1111/j.1530-2415.2009.01195.x

Gross, E, F., and Hardin, C. D. (2007). Implicit and explicit stereotyping of adolescents. *Social Justice Research*, *20*, 140–160.

Han, H. A., Olson, M. A., and Fazio, R. H. (2006). The influence of experimentally created extrapersonal associations on the Implicit Association Test. *Journal of Experimental Social Psychology*, *42*, 259–272.

Hardin, C. D., and Conley, T. D. (2001). A relational approach to cognition: Shared experience and relationship affirmation in social cognition. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton Symposium on the Legacy and Future of Social Cognition* (pp. 3–17). Mahwah, NJ: Erlbaum.

Hardin, C. D., and Higgins, E. T. (1996). Shared reality: How social verification makes the subjective objective. In E. T. Higgins and R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: The interpersonal context* (Vol. 3, pp. 28–84). New York: Guilford.

Hardin, C. D., and Rothman, A. J. (1997). Rendering accessible information relevant: The applicability of everyday life. In R. S. Wyer (Ed.), *The automaticity of everyday life: Advances in social cognition* (Vol. 10, pp. 143–156). Mahwah, NJ: Erlbaum.

Harrison, L. A., Stevens, C. M., Monty, A. N., and Coakley, C. A. (2006). The consequences of stereotype threat on the academic performance of white and non-white lower income college students. *Social Psychology of Education*, *9*, 341–357.

Henry, P. J., and Hardin, C. D. (2006). The contact hypothesis revisited: Status bias in the reduction of implicit prejudice in the United States and Lebanon. *Psychological Science*, *17*, 862–868.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins and A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: Guilford.

Higgins, E. T., Rholes, W. S., and Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*, 141–154.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., and Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*, 1369–1385.

Hugenberg, K., and Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, *14*, 640–643.

Jackman, M. (1994). *The velvet glove: Paternalism and conflict in gender, class, and race relations.* Berkeley, CA: University of California Press.

Jost, J. T., and Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, *33*(1), 1–27. doi:10.1111/j.2044-8309.1994.tb01008.x

Jost, J. T., Banaji, M. R., and Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*, 881–919.

Jost, J. T., and Burgess, D. (2000). Attitudinal ambivalence and the conflict between group and system justification motives in low status groups. *Personality and Social Psychology Bulletin*, *26*, 293–305.

Jost, J. T., Pelham, B. W., and Carvallo, M. R. (2002). Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups. *Journal of Experimental Social Psychology*, *38*(6), 586–602. doi:10.1016/S0022-1031(02)00505-X

Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., and Hardin, C. D. (2009). The existence of implicit prejudice is beyond scientific doubt: A refutation of ideological and methodological objections and executive summary of ten studies no manager should ignore. *Research in Organizational Behavior*, *29*, 39–69.

Jost, J. T., and Thompson, E. P. (2000). Group-based dominance and opposition to equality as independent predictors of self-esteem, ethnocentrism, and social policy attitudes among African Americans and European Americans. *Journal of Experimental Social Psychology*, *36*, 209–232.

Kang, J. and Banaji, M. R. (2006). Fair measures: A behavioral realist revision of "affirmative action." *California Law Review*, *94*, 1063–1118.

Keifer, A. K., and Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility

to stereotype threat. *Journal of Experimental Social Psychology*, *43*, 825–832.

Kim, D-Y. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly*, *66*, 83–96.

Knowles, E. D., Lowery, B. S., and Schaumberg, R. L. (2010). Racial prejudice predicts opposition to Obama and his health care reform plan. *Journal of Experimental Social Psychology*, *46*(2), 420–423. doi:10.1016/j.jesp.2009.10.011

Kressin, N. R., and Petersen, L. A. (2001). Racial differences in the use of invasive cardiovascular procedures: Review of the literature and prescription for future research. *Annual Review of Internal Medicine*, *135*, 352–366.

Lambert, A. J., Payne, B. K., Ramsey, S., and Shaffer, L. M. (2005). On the predictive validity of implicit attitude measures: The moderating effect of perceived group variability. *Journal of Experimental Social Psychology*, *41*, 114–128.

La Pierre, R. (1934). Attitude vs action. *Social Forces*, *13*, 230–237.

Lee, S., Rogge, R. D., and Reis, H. T. (2010). Assessing the seeds of relationship decay. *Psychological Science*, *21*(6), 857–864. doi:10.1177/0956797610371342

Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology*, *71*, 1092–1107.

Lieberman, M. D. (2000). Intuition: A social cognitive neuroscience approach. *Psychological Bulletin*, *126*, 109–137.

Livingston, R. W. (2002). The role of perceived negativity in the moderation of African Americans' implicit and explicit racial attitudes. *Journal of Experimental Social Psychology*, *38*, 405–413.

Lowery, B. S., Hardin, C. D., and Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, *81*, 842–855.

Maison, D., Greenwald, A. G., and Bruin, R. H. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology*, *14*, 405–415.

Marsh, K. L., Johnson, B. L., and Scott-Sheldon, L. A. (2001). Heart versus reason in condom use: Implicit versus explicit attitudinal predictors of sexual behavior. *Zeitschrift fur Experimentelle Psychologie*, *48*, 161–175.

McConnell, A. R., and Liebold, J. M. (2002). Relations between the Implicit Association Test, explicit racial attitudes, and discriminatory behavior. *Journal of Experimental Social Psychology*, *37*, 435–442.

Meyer, D. E., and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.

Mitchell, J. A., Nosek, B. A., and Banaji, M. R. (2003).

Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*, 455–469.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory and Cognition*, *4*, 648–654.

———. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 225–254.

Nock, M. K., and Banaji, M. R. (2007). Prediction of suicide ideation and attempts among adolescents using a brief performance-based test. *Journal of Consulting and Clinical Psychology*, *75*, 707–715.

Nock, M. K., Park, J. M., Finn, C. T., Deliberto, T. L., Dour, H. J., and Banaji, M. R. (2010). Measuring the suicidal mind. *Psychological Science*, *21*(4), 511–517. doi:10.1177/0956797610364762

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, *134*, 565–584.

Nosek, B. A., Banaji, M. R., and Greenwald, A. G. (2002a). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics*, *6*, 101–115.

———. (2002b). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology*, *83*, 44–59.

Nosek, M. A. (1995). Sexual abuse of women with physical disabilities. In T. N. Monga (Ed.), *Physical medicine and rehabilitation state of the art reviews: Sexuality and disability* (pp. 487–502). Philadelphia: Hanley Belfus.

Olson, M. A., and Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*, *12*, 413–417.

———. (2002). Implicit acquisition and manifestation of classically conditioned attitudes. *Social Cognition*, *20*, 89–104.

———. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science*, *14*, 636–639.

———. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*, 653–667.

———. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32*, 421–433.

Orfield, G. (2001). *Schools more separate: Consequences of a decade of resegregation*. Cambridge, MA: Harvard University.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181–192.

Payne, B. K., Shimizu, Y., and Jacoby, L. L. (2005). Mental control and visual illusions: Toward explaining race-biased weapon identifications. *Journal of Experimental Social Psychology*, *41*, 36–47.

Penner, L. A., Dovidio, J. F., West, T. V., Gaertner, S. L., Albrecht, T. L., Dailey, R. K., and Markova, T. (2010). Aversive racism and medical interactions with Black patients: A field study. *Journal of Experimental Social Psychology*, *46*(2), 436–440. doi:10.1016/j.jesp.2009 .11.004

Perdue, C. W., and Gurtman, M. B. (1990). Evidence for the automaticity of ageism. *Journal of Experimental Social Psychology*, *26*, 199–216.

Perugini, M. (2005). Predictive models of implicit and explicit attitudes. *British Journal of Social Psychology*, *44*, 29–45.

Perugini, M. O'Gorman, R., and Prestwich, A. (2007). An ontological test of the IAT: Self-activation can increase predictive validity. *Experimental Psychology*, *54*, 134–147.

Petersen, L. A., Wright, S. M., Peterson, E. D., and Daley, J. (2002). Impact of race on cardiac care and outcomes in veterans with acute myocardial infarction. *Medical Care*, *40*, 186–196.

Phelps, E. A., O'Conner, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., and Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, *12*, 1–10.

Plant, E. A., and Peruche, B. M. (2005). The consequences of race for police officers' responses to criminal suspects. *Psychological Science*, *16*, 180–183.

Plant, E. A., Peruche, B. M., and Butz, D. A. (2005). Eliminating automatic racial bias: Making race non-diagnostic for responses to criminal suspects. *Journal of Experimental Social Psychology*, *41*, 141–156.

Pletcher M. J., Kertesz, S. G., Kohn, M.A., and Gonzales, R. (2008). Trends in opioid prescribing by race/ethnicity for patients seeking care in US emergency departments. *Journal of the American Medical Association*, *299*, 70–78.

Putnam, R. D. (2007). *E Pluribus Unum*: Diversity and community in the twenty-first century, The 2006 Johan Skytte Prize Lecture. *Scandinavian Political Studies*, *30*, 137–174.

Ranganath, K. A., Smith, C. T., and Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement. *Journal of Experimental Social Psychology*, *44*(2), 386–396.

Ratcliff, R., and McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95*, 385–408.

———. (1994). Retrieving information from memory: Spreading activation theories versus compound cue theories. *Psychological Review*, *101*, 177–184.

Richeson, J. A., and Ambady, N. (2003). Effects of situational power on automatic racial prejudice. *Journal of Experimental Social Psychology*, *39*, 177–183.

Richeson, J. A., and Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology*, *40*(3), 417–423. doi:10.1016/j.jesp.2003.09.002

Richeson, J. A., and Shelton, J. N. (2005). Thin slices of racial bias. *Journal of Nonverbal Behavior*, *29*, 75–86.

Richeson, J. A., and Trawalter, S. (2005). Why do interracial interactions impair executive function? A resource depletion account. *Journal of Personality and Social Psychology*, *88*, 934–947.

Roccato, M., and Zogmaister, C. (2010). Predicting the vote through implicit and explicit attitudes: A field research. *Political Psychology*, *31*, 249–274.

Rooth, D-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, *17*(3), 523–534.

Ross, L., and Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* New York: McGraw-Hill.

Rudman, L. A., and Ashmore, R. D. (2007). Discrimination and the Implicit Association Test. *Group Processes and Intergroup Relations*, *10*, 359–372.

Rudman, L. A., Ashmore, R. D., and Gary, M. L. (2001). 'Unlearning' automatic biases: The malleability of implicit stereotypes and prejudice. *Journal of Personality and Social Psychology*, *81*, 856–868.

Rudman, L. A., and Borgida, E. (1995). The afterglow of construct accessibility: The behavioral consequences of priming men to view women as sexual objects. *Journal of Experimental Social Psychology*, *31*, 493–517.

Rudman, L. A., Feinberg, J. M., and Fairchild, K. (2002). Minority members' implicit attitudes: Ingroup bias as a function of ingroup status. *Social Cognition*, *20*, 294–320.

Rudman, L. A., and Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, *57*, 743–762.

Rudman, L. A. and Goodwin, S. A. (2004). Gender differences in automatic ingroup bias: Why do women like women more than men like men? *Journal of Personality and Social Psychology*, *87*, 494–509.

Rudman, L. A., and Heppen, J. (2003). Implicit romantic fantasies and women's interest in personal power: A glass slipper effect? *Personality and Social Psychology Bulletin*, *29*, 1357–1370.

Rudman, L. A., and Lee, M. R. (2002). Implicit and explicit consequences of exposure to violent and misogynous rap music. *Group Processes and Intergroup Relations*, *5*, 133–150.

Rutland, A., Cameron, L., Bennett, L., and Ferrell, J. (2005). Interracial contact and racial constancy: A multi-site study of racial intergroup bias in 3–5 year old Anglo-British children. *Journal of Applied Developmental Psychology*, *26*(6), 699–713. doi:10.1016/j.appdev .2005.08.005

Sagar, H. A., and Schofield, J. W. (1980). Racial and behavioral cues in black and white children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology*, *39*, 590–598.

Sears, D. O., and Henry, P. J. (2005). Over thirty years later: A contemporary look at symbolic racism. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 37, pp. 95–150). San Diego, CA: Academic Press.

Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., and von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology*, *39*(1), 75–82. doi:10.1016/S0022-1031(02)00512-7

Sherman, S. J., Presson, C. J., Chassin, L., Rose, J. S., and Koch, K. (2002). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology*, *22*, 13–39.

Shih, M., Ambady, N., Richeson, J. A., Fujita, K., and Gray, H. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology*, *83*, 638–647.

Shih, M., Pittinsky, T. L., and Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, *10*, 81–84.

Shook, N. J., and Fazio, R. H. (2007). *The effect of interracial versus same-race roommate relationships on attitudes.* Poster presented at the Annual Conference for the Society for Psychology and Social Psychology, Memphis, TN.

Sinclair, S., Hardin, C. D., and Lowery, B. S. (2006). Self-stereotyping in the context of multiple social identities. *Journal of Personality and Social Psychology*, *90*, 529–542.

Sinclair, S., Lowery, B. S., and Dunn, E. (2005). The relationship between parental racial attitudes and children's implicit prejudice. *Journal of Experimental Social Psychology*, *14*, 283–289.

Sinclair, S., Lowery, B. S., Hardin, C. D., and Colangelo, A. (2005). Social tuning of automatic racial attitudes: The role of affiliative motivation. *Journal of Personality and Social Psychology*, *89*, 583–592.

Skinner, B. F. (1953). *Science and human behavior.* New York: Macmillan.

Sniderman, P. M., and Carmines, E. G. (1997). *Reaching beyond race.* Cambridge, MA: Harvard University Press.

Spalding, L. R., and Hardin, C. D. (1999). Unconscious unease and self-handicapping: Behavioral consequences of individual differences in implicit and explicit self-esteem. *Psychological Science*, *10*, 535–539.

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*, 4–28.

Srull, T. K., and Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, *38*, 841–856.

Steele, C. M., and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797–811.

Steffens, M. C. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology*, *51*, 165–179.

Swanson, J. E., Rudman, L. A., and Greenwald, A. G. (2001). Using the Implicit Association Test to investigate attitude-behavior consistency for stigmatized behavior. *Cognition and Emotion*, *15*, 207–230.

Sylvestre, D. L., Litwin, A. H., Clements, B. J., and Gourevitch, M. N. (2005). The impact of barriers to hepatitis C virus treatment in recovering heroin users maintained on methadone. *Journal of Substance Abuse Treatment*, *29*, 159–165.

Tam, T., Hewstone, M., Cairns, E., Tausch, N., Maio, G., and Kenworthy, J. B. (2007). The impact of intergroup emotions on forgiveness in Northern Ireland. *Group Processes and Intergroup Relations*, *10*, 119–135.

Tam, T., Hewstone, M., Harwood, J., Voci, A., and Kenworthy, J. (2006). Intergroup contact and grandparent-grandchild communication: The effects of self-disclosure on implicit and explicit biases against older people. *Group Processes and Intergroup Relations*, *9*, 413–430.

Teachman, B. A., Marker, C. D., and Smith-Janik, S. B. (2008). Automatic associations and panic disorder: Trajectories of change over the course of treatment. *Journal of Consulting and Clinical Psychology*, *76*(6), 988–1002.

Teachman, B. A., Smith-Janik, S. B., and Saporito, J. (2007). Information processing biases and panic disorder: Relationships among cognitive and symptom measures. *Behaviour Research and Therapy*, *45*, 1791–1811.

Teachman, B. A., and Woody, S. R. (2003). Automatic processing in spider phobia: Implicit fear associations over the course of treatment. *Journal of Abnormal Psychology*, *112*, 100–109.

Tetlock, P. E., and Mitchell, G. (in press). Unconscious prejudice and accountability systems: What must organizations do to check implicit bias? *Research in Organizational Behavior.*

Turner, R. N., Hewstone, M., and Voci, A. (2007). Reducing explicit and implicit outgroup prejudice via direct and extended contact: The mediating role of self-disclosure and intergroup anxiety. *Journal of Personality and Social Psychology*, *93*(3), 369–388. doi:10.1037/0022-3514.93.3.369

U. S. Department of Justice. (2001). *Policing and homicide, 1976–98: Justifiable homicide by police, police officers murdered by felons* (NCJ 180987). Washington, DC: Bureau of Justice Statistics.

van Breukelen, W., van der List, R., and Steensma, H. (2004). Voluntary employee turnover: Combining variables from the "traditional" turnover literature

with the theory of planned behavior. *Journal of Organizational Behavior*, *25*, 893–914.

von Hippel, W., Brener, L., and von Hippel, C. (2008). Implicit prejudice toward injecting drug users predicts intentions to change jobs among drug and alcohol nurses. *Psychological Science*, *19*, 7–11.

Wegener, D. T., and Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 141–208). San Diego, CA: Academic Press.

Weirs, R. W., Woerden, N. V., Smulders, F. T., and de Jong P. T. (2002). Implicit and explicit alcohol-related cognitions in heavy and light drinkers. *Journal of Abnormal Psychology*, *111*, 648–658.

Wittenbrink, B., Judd, C. M., and Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship to questionnaire measures. *Journal of Personality and Social Psychology*, *72*, 262–274.

Wittenbrink, B., Judd, C. M., and Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*(5), 815–827. doi:10.1037/0022-3514.81.5.815

Wittenbrink, B., and Schwarz, N. (Eds.) (2007). *Implicit measures of attitudes*. New York: Guilford.

Word, C. O., Zanna, M. P., and Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology*, *10*, 109–120.

Wraga, M., Helt, M., Jacobs, E., and Sullivan, K. (2007). Neural basis of stereotype-induced shifts in women's mental rotation performance. *Social Cognitive and Affective Neuroscience*, *2*, 12–19.

Ziegert, J. C., and Hanges, P. J. (2005). Employment discrimination: The role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology*, *90*, 553 –562.

Biases in Interracial Interactions

Implications for Social Policy

J. NICOLE SHELTON

JENNIFER A. RICHESON

JOHN F. DOVIDIO

The *Brown v. Board of Education* decision in 1954 and the Civil Rights Act of 1964 were monumental policy decisions that changed the landscape of race relations in the United States. Before the implementation of these policies, ethnic minorities and Whites had very little contact with one another, primarily because ethnic minorities were not allowed to be in the same settings with Whites, including attending the same schools, working in the same place of employment, living in the same neighborhoods, and even riding in the same sections of buses and eating in the same sections of restaurants. When contact between the groups occurred, it was often fraught with extreme hostility, fear, and anxiety. Without a doubt, the policy decisions of 1954 and 1964 against segregation and discrimination increased the opportunity for people to engage in contact with members outside of their racial group. In addition, these policy decisions paved the way for improvements in social norms toward ethnic minorities, such that it was unacceptable to publicly express negative racial beliefs or behave in a discriminatory manner toward them. The change in social norms eventually improved individuals' private attitudes and behaviors. Indeed, there has been a substantial increase in Whites' endorsement of racial equality and integration over the past fifty-plus years (Bobo, 2001).

Although federal laws and organizational policies have been developed and have been relatively effective in reducing blatant forms of bias (Fiske and Krieger; Hardin and Banaji, this volume), it is more challenging to create laws and policies to reduce the subtle bias that is often present in everyday interracial interactions. For example, a law cannot be created that prohibits Whites from displaying negative nonverbal behaviors toward African Americans during daily interactions. Yet, subtle negative behavior and signals can have adverse effects on the performance and ambitions of African Americans (Purdie-Vaughns et al., 2008; Salvatore and Shelton, 2007). Moreover, efforts to demand that individuals comply with egalitarian social norms can provoke hostility and have the unintended consequence of worsening the problems (Plant and Devine, 2001). Given the problems that both blatant and subtle bias create, it is important to understand how both types influence everyday interracial interactions and how policies may help improve the quality of interracial interactions and ultimately reduce racial bias.

In this chapter, we explore how and why racial bias systematically influences daily interracial interactions across three contexts: (a) residential spaces on college campuses, (b) health care provider-patient dyads in medical settings, and (c) employee-employer relations in the workplace. We focus our attention specifically on these contexts, in part, because the landmark policy decisions of 1954 and 1964 opened the doors for increased opportunities for individuals to engage in contact across the racial divide in these three areas. Moreover, we selected these contexts because they are venues into which most people will enter at some point during their lives. With respect to residential experiences in colleges, we are aware, of course, that not everyone attends college. However, given that policies about the American educational system played such a profound role in creating spaces for intergroup contact, it seems essential to explore the dynamics of contact in this setting. In addition, some of the most influential work on contact theory occurred in residential areas and revealed that contact was related to improved intergroup attitudes (e.g., Deutsch and Collins, 1951; Wilner, Walkley, and Cook, 1955). Moreover, studying roommate relationships on college campuses may offer insight into ways to improve the academic outcomes of students (see Cohen and Garcia, this volume for policies about the racial achievement gap).

Regardless of educational level, the majority of people in the United States will interact with some form of medical-care provider and will obtain employment at some point in their lives. Racial bias in these contexts has the potential to inhibit successful life outcomes. Health-care providers' racial biases, for example, have been posited to contribute to racial disparities in health outcomes (Dovidio et al., 2008). The Institute of Medicine's 2003 report on unequal medical treatment acknowledge that such biases, if they exist, undermine the high ethical standards that medical professionals are accountable for upholding

(Smedley, Stith, and Nelson, 2003).[1] Describing how racial biases interfere with the behaviors of a group of well-meaning people who have taken an oath to treat all individuals equally highlights the insidious nature of racial bias. Understanding when and how racial biases may influence racial disparities in health care outcomes may begin to shed light on policies that could be developed to improve the life expectancy of ethnic minorities.

Finally, we focus on the workplace because, although there are explicit federal laws and sanctions against blatant racial bias in the workplace, many organizational-level practices are such that subtle bias remains a problematic force. These practices trickle down to influence the ways in which Whites and ethnic minorities interact with one another in their daily lives while at work. Taken together, these three contexts are common avenues for interracial contact to occur and are ones in which policies and regulations have been put forth, or could be established, to improve the quality of Whites' and ethnic minorities' life experiences.

The chapter is divided into three primary sections. In the first, we provide an overview of the literature on contemporary intergroup bias, particularly with respect to race in the United States. In the second, we discuss the processes associated with the interplay between racial attitudes and interracial contact. And in the third, we discuss the implications of these processes for the aforementioned three contexts, paying special attention to how policies in these contexts may shape individuals' experiences during interracial encounters and how knowing individuals' experiences may offer ideas about policy decisions. We pay particular attention to the fact that racial biases may have different consequences, in some cases completely opposite effects, for Whites and ethnic minorities during their interactions in these settings. These divergent experiences pose a major challenge for policy: policies must be tailored in such a way that an improvement in the lives of one group does not cause harm in the lives of the other. In essence, we explore how policy decisions shape the psychology of interracial interactions and how the psychology of interracial interactions may shape policies established to create harmonious interracial relations. Because we cover a lot of ground by describing interracial contact across three settings, our review of the literature is not meant to be exhaustive. Instead, we sample a few classic and contemporary articles that highlight the issues of concern in the best light. In addition, we focus primarily on race relations in the United States within the context of Black-White relations, which have historically been central to the development of social policy defining

intergroup relations in the United States. Finally, we offer a synthesis of common practices based on intergroup contact research that policy makers use to maximize the benefits of diversity across multiple settings.

Intergroup Bias

Intergroup bias, a pervasive and arguably universal phenomenon within and across many cultures (Sidanius and Pratto, 1999), stems from processes associated with prejudice and stereotyping. Prejudice reflects a general negative evaluation of a group, whereas stereotyping reflects the association of specific traits to a group. Prejudice and stereotypes often lead to discrimination, which is the unjustified group-based difference in behavior that gives one group an advantage over others. Perhaps intergroup bias is a pervasive phenomenon because there are several normal processes that allow people to navigate a complex environment that predispose them to developing intergroup prejudices. For example, the ability to sort people, spontaneously and with minimum effort and awareness, into meaningful categories is a universal facet of human perception essential for efficient functioning (Bodenhausen, Todd, and Becker, 2007). Given the importance of the self in social perception, social categorization further involves a basic distinction between the group containing the self (in-group) and other groups (out-groups)—or between the "we's" and the "they's" (Turner et al., 1987). The recognition of different group memberships shapes social perception, affect, cognition, and behavior in ways that systematically produce intergroup biases (Gaertner and Dovidio, 2000). If, when, and how bias is manifested, however, depends upon cultural norms, individual motivation, the historical relations between groups, and the immediate circumstances (Crandall and Eshleman, 2003). In societies that place high value on egalitarianism, going as far as establishing laws to promote equality, intergroup biases often take the form of subtle, rather than blatant, prejudice (Gaertner and Dovidio, 1986).

The discrepancy between the ideal of egalitarianism and the psychological forces that promote racial bias has been posited as a critical factor leading to the development of subtle forms of racial bias. Whereas the traditional form of racial bias represented the overt expression of dislike and hostility, as well as the endorsement of negative cultural stereotypes, contemporary forms of racial bias involve more complex dynamics and typically more subtle expressions of bias. This is evident in research within a framework of *aversive racism* (Dovidio and Gaertner, 2004), which