IMPLICIT BIAS IN THE COURTROOM

Jerry Kang

Judge Mark Bennett

Devon Carbado

Pam Casey

Nilanjana Dasgupta

David Faigman

Rachel Godsil

Anthony Greenwald

Justin Levinson

Jennifer Mnookin

ABSTRACT

AUTHORS

Jerry Kang is Professor of Law at UCLA School of Law; Professor of Asian American Studies (by courtesy); Korea Times-Hankook Ilbo Chair in Korean American Studies. kang@law.ucla.edu, http://jerrykang.net. Judge Mark Bennett is a U.S. district judge in the Northern District of Iowa. Devon Carbado is Professor of Law at UCLA School of Law. Pam Casey is Principal Court Research Consultant of the National Center for State Courts. Nilanjana Dasgupta is Associate Professor of Psychology at the University of Massachusetts at Amherst. David Faigman is John F. Digardi Distinguished Professor of Law, University of California Hastings College of the Law; Director, UCSF/UC Hastings Consortium on Law, Science and Health Policy; Professor, University of California, San Francisco, School of Medicine, Dept. of Psychiatry. Rachel Godsil is Eleanor Bontecou Professor of Law at Seton Hall University School of Law. Anthony Greenwald is Professor of Psychology at the University of Washington. Justin Levinson is Associate Professor of Law and Director of Culture and Jury Project at the University of Hawai'i William S. Richardson School of Law. Jennifer Mnookin is Professor of Law at UCLA School of Law.

ABSTRACT

Given the substantial and growing scientific literature on implicit bias, the time has now come to confront a critical question: *What, if anything, should we do about implicit bias in the courtroom?* The author team comprises legal academics, scientists, researchers, and even a sitting federal judge who seek to answer this question in accordance with "behavioral realism." The Article first provides a succinct scientific introduction to implicit bias, with some important theoretical clarifications that distinguish between explicit, implicit, and structural forms of bias. Next, the article applies the science to two trajectories of bias relevant to the courtroom. One story follows a criminal defendant path; the other story follows a civil employment discrimination path. This application involves not only a focused scientific review but also a step-by-step examination of how criminal and civil trials proceed. Finally, the Article examines various concrete intervention strategies to counter implicit biases for key players in the justice system, such as the judge and jury.

TABLE OF CONTENTS

INTRODUCTION

The problems of overt discrimination have received enormous attention from lawyers, judges, academics, and policymakers. While explicit sexism, racism, and other forms of bias persist, they have become less prominent and public over the past century. But explicit bias and overt discrimination are only part of the problem. Also important, and likely more pervasive, are questions surrounding implicit bias—attitudes or stereotypes that affect our understanding, decisionmaking, and behavior, without our even realizing it.

How prevalent and significant are these implicit, unintentional biases? To answer these questions, people have historically relied on their gut instincts and personal experiences, which did not produce much consensus. Over the past two decades, however, social cognitive psychologists have discovered novel ways to measure the existence and impact of implicit biases—without relying on mere "common sense." Using experimental methods in laboratory and field studies, researchers have provided convincing evidence that implicit biases exist, are pervasive, large in magnitude, and have real-world effects. These fascinating discoveries, which have migrated from the science journals into the law reviews and even popular discourse, are now reshaping the law's fundamental understandings of discrimination and fairness.

Given the substantial and growing scientific literature on implicit bias, the time has now come to confront a critical question: *What, if anything, should we do about implicit bias in the courtroom?* In other words, how concerned should we be that judges, advocates, litigants, and jurors are coming to the table with implicit biases that influence how they interpret evidence, understand the facts, parse legal principles, and make judgment calls? In what circumstances are these risks most acute? Are there practical ways to reduce the effects of implicit biases? To what extent can awareness of these biases mitigate against their impact? What other 'debiasing' strategies might work? In other words, in what way—if at all— should the courts respond to a better model of human decisionmaking that the mind sciences are providing?

We are a team of legal academics, scientists, researchers, and even a sitting federal judge[1] who seek to answer these difficult questions in accordance with "behavioral realism."[2] Our general goal is to educate those

---

1.   Judge Mark W. Bennett, a coauthor of this article, is a United States District Court Judge in the Northern District of Iowa.

2.   Behavioral realism is a school of thought that asks the law to account for more accurate models of human cognition and behavior. *See, e.g.,* Jerry Kang & Kristin Lane, *Seeing Through Colorblindness: Implicit Bias and the Law*, 58 UCLA L. REV. 465, 490 (2010); Linda Hamilton Krieger & Susan T. Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 CALIF. L. REV. 997, 997–1008 (2006). Jon Hanson and his coauthors have advanced similar approaches under the names of "critical realism,"

in the legal profession who are unfamiliar with implicit bias and its consequences. To do so, we provide a current summary of the underlying science, contextualized to criminal and civil litigation processes that lead up to and crescendo in the courtroom. This involves not only a focused scientific review but also a step-by-step examination of how criminal and civil trials proceed, followed by suggestions designed to address the harms. We seek to be useful to legal practitioners of good faith, including judges, who conclude that implicit bias is a problem (one among many) but don't know quite what to do about it. While we aim to provide useful and realistic strategies for those judges already persuaded that implicit bias is a legitimate concern, we also hope to provoke those who know less about it, or are more skeptical of its relevance, to give these issues thoughtful consideration.

We are obviously not a random sample of researchers and practitioners; thus, we cannot claim any representative status. That said, the author team represents a broad array of experience, expertise, methodology, and viewpoints. In authoring this paper, the team engaged in careful deliberations, across topics of both consensus and dissensus.[3] We did not entirely agree on how to frame questions in this field or how to answer them. That said, we stand collectively behind what we have written. We also believe the final work product reveals the benefits of such cross-disciplinary and cross-professional collaboration.

Part I provides a succinct scientific introduction to implicit bias, with some important theoretical clarifications. Often the science can seem too abstract, especially to nonprofessional scientists. As a corrective, in Part II, we apply the science to two trajectories of bias relevant to the courtroom. One story follows a criminal defendant path; the other story follows a civil employment discrimination path. Part III examines different intervention strategies to counter implicit biases for key players in the justice system, such as the judge and jury.

"situationism," and the "law and mind sciences." *See* Adam Benforado, *Frames of Injustice: The Bias We Overlook*, 85 Ind. L.J. 1333, 1339 n.28 (2010) (listing papers).

      3.   This paper arose out of the second symposium of PULSE: Program on Understanding Law, Science, and Evidence at UCLA School of Law (March 3–4, 2011). We brought together leading scientists (including Anthony Greenwald, the inventor of the Implicit Association Test), federal and state judges, applied researchers, and legal academics to explore the scientific state-of-the-art regarding implicit bias research and to examine the various institutional responses to date. The Symposium also raised possibilities and complications, ranging from the theoretical to practical, from the legal to the scientific. After a day of public presentations, the author team met in closed session for a full day to craft the outlines of this paper. Judge Michael Linfield of the Los Angeles Superior Court as well as Jeff Rachlinski, Professor of Law at Cornell Law School, participated in the symposium but could not join the author team. Their absence should not be viewed as either agreement or disagreement with the contents of the Article.

I.   Implicit Biases

A.   Empirical Introduction

Over the past thirty years, cognitive and social psychologists have demonstrated that human beings think and act in ways that are often not rational. We suffer from a long litany of biases, most of them having nothing to do with gender, ethnicity, or race. For example, we have an oddly stubborn tendency to "anchor" to numbers, judgments, or assessments that we have been exposed to and use them as a starting point for future judgments—even if those anchors are objectively wrong.[4] We exhibit an "endowment effect," with irrational attachments to arbitrary initial distributions of property, rights, and grants of other entitlements.[5] We suffer from "hindsight bias" and believe that what turns out to be the case today should have been easily foreseen yesterday.[6] The list of empirically-revealed biases goes on and on. Indeed, many legal academics have become so familiar with such heuristics and biases that they refer to them in their analyses as casually as they refer to economic concepts such as "transaction costs."[7]

One type of bias is driven by attitudes and stereotypes that we have about social categories, such as genders and races. An *attitude* is an association between some concept (in this case a social group) and an evaluative valence, either positive or negative.[8] A *stereotype* is an association between a concept (again, in this case a social group) and a trait.[9] Although interconnected, attitudes and stereotypes should be distinguished because a positive attitude does not foreclose negative stereotypes and vice versa. For instance, one might have a positive overall attitude toward African Americans and yet still associate them with weapons. Or, one might have a positive stereotype of Asian Americans as mathematically-able but still have an overall negative attitude towards them.

---

4.   *See* Jon D. Hanson & Douglas A. Kysar, *Taking Behavioralism Seriously: The Problem of Market Manipulation*, 74 N.Y.U. L. Rev. 630, 667 (1999) (describing anchoring).

5.   *See generally* Russell Korobkin, *The Endowment Effect and Legal Analysis*, 97 97 Nw. U. L. Rev. 1227 (2003).

6.   *See generally* Daniel Kahneman, Thinking, Fast and Slow (2011); Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 U. Chi. L. Rev. 571 (1998).

7.   *See, e.g.*, Russell B. Korobkin & Thomas S. Ulen, *Law and Behavioral Science: Removing The Rationality Assumption From Law and Economics*, 88 Cal. L. Rev. 1051 (2000); Donald C. Langevoort, *Behavioral Theories of Judgment and Decision Making in Legal Scholarship: A Literature Review*, 51 Vand. L. Rev. 1499 (1998).

8.   In both common and expert usage, sometimes the word "prejudice" is used to describe a negative attitude, especially when it is strong in magnitude.

9.   If the association is nearly perfect in that almost every member of that social group has that trait, then we think of the trait less as a stereotype and more as a defining attribute. Typically, when we use the word "stereotype," the correlation between social group and trait is far from perfect. *See* Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 Calif. L. Rev. 945, 949 (2006).

The conventional wisdom has been that these social cognitions—attitudes and stereotypes about social groups—are explicit, in the sense that they are both consciously accessible through introspection *and* endorsed by the person as appropriate. Indeed, this understanding has shaped much of current antidiscrimination law. The conventional wisdom is also that the social cognitions that individuals hold are relatively stable in the sense that they operate in the same way over time and across different situations.

However, recent findings in the mind sciences, especially implicit social cognition (ISC),[10] have undermined these conventional beliefs. As detailed below, attitudes and stereotypes may also be implicit, in the sense that they are not consciously accessible through introspection. Accordingly, their impact on a person's decisionmaking and behaviors does not depend on that person's awareness of possessing these attitudes or stereotypes. In this way, they can function automatically, including in ways that the person would not endorse as appropriate if he or she did have conscious awareness. Moreover, evidence suggests that stereotypes and especially attitudes may be highly reactive to salient cues in the environment, and thus relatively stable within a person only to the extent that that person remains in relatively similar environmental settings.

How have mind scientists discovered such findings on matters so latent or implicit? They have done so by innovating new techniques that allow the measure of implicit attitudes and stereotypes that by definition can't be reliably self-reported. Some of these measures involve subliminal priming and other not consciously-detected treatments within an experimental setting. Other instruments use reaction time differences between two types of tasks—one that seems consistent with some bias, the other inconsistent—as in the Implicit Association Test (IAT).[11]

The well-known IAT is a sorting task that measures time differences between schema-consistent pairings and schema-inconsistent pairings of concepts, as represented by words or pictures. For example, suppose we want to test whether there is an implicit stereotype associating African Americans with weapons. In a schema-consistent run, the participant is instructed to hit the same response key when she sees a White face or a

---

10.   Implicit social cognition (ISC) is a field of psychology that examines the mental processes that affect social judgments but operate without conscious awareness or conscious control. *See generally* Kristin A. Lane, Jerry Kang & Mahzarin Banaji, *Implicit Social Cognition and Law*, 3 Ann. Rev. Law & Soc. Sci. 19.1 (2007). The term was first used and defined by Anthony Greenwald and Mahzarin Banaji. *See* Anthony G. Greenwald & Mahzarin R. Banaji, *Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes*, 102 Psychol. Rev. 4 (1995).

11.   *See* Anthony G. Greenwald et al., *Measuring Individual Differences in Implicit Cognition: The Implicit Association Test*, 74 J. Personality & Soc. Psychol. 1464, 1464–66 (1998) (introducing the Implicit Association Test (IAT)). For more information on the IAT, see Brian A. Nosek, Anthony G. Greenwald, & Mahzarin R. Banaji, *The Implicit Association Test at Age 7: A Methodological and Conceptual Review*, in Automatic Processes In Social Thinking and Behavior 265 (John A. Bargh ed., 2007).

harmless object, and another response key when she sees an African American face or a weapon. Notice that White and harmless item are on the same key; African American and weapon are on the other same key. Most people perform this task quickly.

In a schema-inconsistent run, we reverse the pairings. White and weapon are on the same key; African American and harmless item are on the same key. Most people perform this task more slowly. Of course the order in which these tasks are presented is always systematically varied in order to ensure that the speed of people's responses is not affected by practice. The time differential between these runs is defined as the implicit association effect and is statistically processed into standard units called an IAT D score.[12]

Through the IAT, social psychologists from hundreds of laboratories have collected enormous amounts of data[13] on reaction-time measures of "implicit biases," by which we mean implicit attitudes and implicit stereotypes. According to these measures, implicit bias is pervasive (widely held), large in magnitude (as compared to standardized measures of explicit bias),[14] dissociated from explicit biases (which suggests that explicit biases and implicit biases, while related, are separate mental constructs),[15] and predicts certain kinds of real-world behavior.[16] What policymakers are now keen to understand are the size and scope of these behavioral effects and how to counter them—by altering the implicit biases themselves and by implementing strategies to attenuate their effect.

Useful and current summaries of the scientific evidence can be found in both the legal and psychological literatures. For example, in the last volume of this law review, Jerry Kang and Kristin Lane provided a summary of the evidence that we are not perceptually, cognitively, or behav-

---

12.   This D score, which ranges from –2.0 to 2.0, is a standardized score (computed by dividing the IAT effect by the standard deviations of the participants' latencies pooled across schema-consistent and -inconsistent conditions). *See, e.g.*, Anthony Greenwald et al., *Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm*, 85 J. Personality. & Soc. Psychol. 197 (2003).

13.   The most prominent dataset is collected at Project Implicit, http://projectimplicit .org (last visited Jan. 6, 2012) (providing free online tests of automatic associations). For a broad analysis of this dataset, see Brian A. Nosek et al., *Pervasiveness and Correlates of Implicit Attitudes and Stereotypes*, 18 Eur. Rev. Soc. Psychol. 1 (2007).

14.Cohen's d is a standardized unit of the size of a statistical effect. By convention, social scientists mark 0.20, 0.50, and 0.80 as small, medium, and large effect sizes. The IAT effect, as measured in Cohen's d, on various stereotypes and attitudes range from medium to large. *See* Kang & Lane, *supra* note 2, at 474 n.35 (discussing data from Project Implicit). Moreover, the effect sizes of implicit bias against social groups are frequently larger than the effect sizes produced by explicit bias measures. *See id.* at 474-75 tbl. 1.

15.   *See* Anthony G. Greenwald & Brian A. Nosek, *Attitudinal Dissociation: What Does It Mean?*, *in* Attitudes: Insights From the New Implicit Measures 65 (R.E. Petty & P. Brinol, Eds. 2008).

16.*See* Kang & Lane, *supra* note 2, at 481-90 (discussing evidence of biased behavior in perceiving smiles, responding to threats, body language, and screening resumes).

iorally colorblind.[17] Justin Levinson and Danielle Young have summarized studies focusing on jury decisionmaking.[18] In the psychology journals, John Jost and colleagues responded to sharp criticism[19] that the IAT studies lacked real-world consequences by providing a qualitative review of the literature, including ten studies that no manager should ignore. Further, they explained how the findings are entirely consistent with the major tenets of twentieth century social cognitive psychology.[20] In a quantitative review, Anthony Greenwald conducted a meta-analysis of IAT studies—which quantitatively synthesizes all the relevant scientific findings—and found that implicit attitudes as measured by the IAT predicted certain types of behavior, such as anti-Black discrimination or intergroup discrimination, substantially better than explicit bias measures.[21]

Instead of duplicating these summaries, we offer research findings that are specific to the topic at hand: implicit bias leading up to and in the courtroom. To do so, we chart out two case trajectories—one criminal, the other civil. That synthesis appears in Part II.

B.   Theoretical Clarification

But before we leave our introduction to implicit bias, we seek to make some theoretical clarifications on the relationships between explicit, implicit, and structural processes that are all involved in producing unfairness in the courtroom. We do so because the legal literature has flagged this as an important issue.[22] In addition, a competent diagnosis of unfairness in the courtroom requires disentangling these various processes. For instance, if the end is to counter discrimination caused by, say, *explicit* bias, it may be ineffective to adopt means that are better tailored to respond to *implicit* bias, and vice versa.

We start by clarifying terms. To repeat, "explicit" biases are attitudes and stereotypes that are consciously accessible through introspection *and* endorsed as appropriate. If no social norm against these biases

---

17.   *See* Kang & Lane, *supra* note 2, at 473–90. *See also* David L. Faigman, Nilanjana Dasgupta & Cecilia L. Ridgeway, *A Matter of Fit: The Law of Discrimination and the Science of Implicit Bias*, 59 HASTINGS L. J. 1389 (2008).

18.   *See* Justin D. Levinson & Danielle Young, *Different Shades of Bias: Skin Tone, Implicit Racial Bias, and Judgments of Ambiguous Evidence*, 112 W. VA. L. REV. 307, 319–26 (2010).

19.   *See, e.g.*, Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 OHIO ST. L.J. 1023, 1108–10 (2006).

20.   *See, e.g.*, John T. Jost et al., *The Existence of Implicit Prejudice Is Beyond Reasonable Doubt: A Refutation of Ideological and Methodological Objections and Executive Summary of Ten Studies That No Manager Should Ignore* , 29 RES. ORG. BEHAV. 39, 41 (2009).

21.   *See* Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity*, 97 J. PERSONALITY & SOC. PSYCHOL. 17, 19–20 (2009). Implicit attitude scores predicted behavior in this domain at an average correlation of r=.24, whereas explicit attitude scores had correlations at an average of r=.12. *See id.* at 73, tbl.3.

22.   *See generally* Stephen M. Rich, *Against Prejudice*, 80 GEO. WASH. L. REV. 1 (2011); Ralph Richard Banks & Richard Thompson Ford, *(How) Does Unconscious Bias Matter?: Law, Politics, and Racial Inequality*, 58 EMORY L.J. 1053 (2009).

exist within a given context, a person will freely broadcast them to others. But if such a norm exists, then explicit biases can be "concealed" to manage the impressions that others have of us. By contrast, "implicit" biases are attitudes and stereotypes that are not consciously accessible through introspection. If we find out that we have them, we may indeed reject them as inappropriate.

Above, we used "explicit" and "implicit" as adjectives to describe mental constructs—attitudes and stereotypes. Readers should recognize that these adjectives can also apply to research procedures or instruments. An "explicit" instrument asks the respondent for a direct self-report with no attempt by researchers to disguise the mental construct that is being measured. An example is a straightforward survey question. No instrument perfectly measures a mental construct. In fact, one can often easily conceal one's "explicit" bias measured through an "explicit" instrument. In this way, an explicit instrument can poorly measure an explicit bias, as the test subject may choose not to be candid about the beliefs or attitudes at issue.

By contrast, an "implicit" instrument does not depend on the respondent's conscious knowledge of the mental constructs that the researcher is inferring from the measure. An example is a reaction time measure, such as the IAT. This does not necessarily mean that the respondent is unaware that the IAT is measuring bias. It also does not mean that the respondent is actually unaware that he or she has implicit biases, for example because she has taken an IAT before or is generally aware of the research literature. To repeat, no instrument perfectly measures any mental construct, and this remains true for implicit instruments. One might, for instance, try to "conceal" implicit bias measured through an implicit instrument, but such faking is often much harder than faking explicit bias measured by an explicit instrument.[23]

Finally, besides explicit and implicit biases, another set of processes that produce unfairness in the courtroom can be called "structural." Other names include "institutional" or "societal." These processes can lock in past inequalities, reproduce them, and indeed exacerbate them even without formally treating persons worse simply because of attitudes and stereotypes about the groups to which they belong.[24] In other words, structural bias can produce unfairness even though no single individual is

---

23. *See, e.g.*, Do-Yeong Kim, *Voluntary Controllability of the Implicit Association Test (IAT)*, 66 Soc. Psychol. Q. 83, 95–96 (2003).

24. *See, e.g.*, john a. powell, *Structural Racism: Building Upon Insights of John Calmore*, 86 N.C. L. Rev. 791, 795–800 (2008) (adopting a systems approach to describe structured racialization); Michelle Adams, *Intergroup Rivalry, Anti-Competitive Conduct and Affirmative Action*, 82 B.U. L. Rev. 1089, 1117–22 (2002) (applying lock-in theory to explain the inequalities between Blacks and Whites in education, housing, and employment); Daria Roithmayr, *Barriers to Entry: A Market Lock-In Model of Discrimination*, 86 Va. L. Rev. 727, 743–48 (2000) (describing lock-in theory, drawing on antitrust law and concepts).

being treated worse right now on the basis of his or her membership in a particular social category.

Because thinking through biases with respect to human beings evokes so much potential emotional resistance, sometimes it's easier to apply them to something less fraught than gender, race, religion, and so on. So, consider a vegetarian's biases against meat. He has a negative attitude (that is, "prejudice") toward meat. He also believes that eating meat is bad for his health (a "stereotype"). He is aware of this attitude and stereotype. He also endorses them as appropriate. In other words, he feels that it's okay to have a negative reaction to meat. He also believes it accurate enough to believe that meat is generally bad for human health, and that there's no reason to avoid behaving in accordance with this belief. These are *explicit biases*.

Now, if this vegetarian is running for political office and campaigning in a region famous for barbecue, he will probably keep his views to himself. He could, for example, avoid showing disgust on his face or making critical comments when a plate of ribs is placed in front of him. Indeed, he might even take a bite and compliment the cook. This is an example of *concealed bias* (explicit bias that is hidden for impression management reasons).

Consider, by contrast, another vegetarian who has recently converted for environmental reasons. She proclaims explicitly and sincerely a negative attitude toward meat. But it may well be that she has an implicit attitude that is still slightly positive. Suppose that she grew up enjoying weekend barbecues with family and friends, or still likes the taste of steak, or first learned to cook by making roasts. Whatever the sources and causes, she may still have an implicitly positive attitude toward meat. This is an *implicit bias*.

Now, finally consider some eating decision that she has to make at a local strip mall. She can buy a salad for $10.00 or a cheeseburger for $3.00. Unfortunately, she has only $5.00 to spare and must eat. Neither explicit nor implicit biases much explain her decision to buy the cheeseburger. She simply lacks the funds to buy the salad, and her need to eat trumps her desire to avoid meat. The decision wasn't driven principally by an attitude or stereotype, explicit or implicit, but by the price. But what if a careful historical, economic, political, and cultural analysis revealed multifarious subsidies, political kickbacks, historical contingencies, and economies of scale that accumulated in mutually reinforcing ways to price the salad much higher than the cheeseburger? These various forces could make it more "rational" for consumers to eat cheeseburgers. This would be an example of *structural* bias in favor of meat.

We disentangle these various mechanisms—explicit attitudes and stereotypes (sometimes concealed, sometimes revealed), implicit attitudes and stereotypes, and structural forces—because they pose different threats to fairness everywhere, including the courtroom. For instance, the threat

to fairness posed by jurors with explicit negative attitudes toward Muslims but who conceal their prejudice to stay on the jury is quite different from the threat posed by jurors who perceive themselves as nonbiased but who nevertheless hold negative implicit stereotypes about Muslims. Where appropriate, we explain how certain studies provide evidence of one type of bias or the other. In addition, we want to underscore that these various mechanisms—explicit bias, implicit bias, and structural forces—are not mutually exclusive.[25] To the contrary, they may often be mutually rein-forcing. In focusing on implicit bias in the courtroom, we do not mean to suggest that implicit bias is the only or most important problem, or that explicit bias (revealed or concealed) and structural forces are unimportant or insignificant.[26]

II.   Two Trajectories

A.   The Criminal Path

Consider, for example, some of the crucial milestones in a criminal case flowing to trial. First, on the basis of a crime report, the police inves-tigate particular neighborhoods and persons of interest and ultimately ar-rest a suspect. Second, the prosecutor decides to charge the suspect with a particular crime. Third, the judge makes decisions about bail and pre-trial detention. Fourth, the defendant decides whether to accept a plea bargain, after consulting his defense attorney, often a public defender or court-appointed private counsel. Fifth, if the case goes to trial, the judge manag-es the proceedings while the jury decides whether the defendant is guilty. Finally, if convicted, the defendant must be sentenced. At each of these stages,[27] implicit biases can have an important impact. To maintain a manageable scope of analysis, we focus on the police encounter, charge and plea bargain, trial, and sentencing.

1.   Police Encounter

Blackness and criminality. If we implicitly associate certain groups, such as African Americans, with certain attributes, such as criminality, then it should not be surprising that police may behave in a manner con-

---

25.   *See, e.g.,* Glenn C. Loury, The Anatomy of Racial Inequality 23–30 (2002) (discussing self-reinforcing stereotypes); john a. powell & Rachel Godsil, *Implicit Bias Insights as Preconditions to Structural Change*, 20 Poverty & Race (2011) (explaining why "implicit bias insights are crucial to addressing the substantive inequalities that result from structural racialization").

26.   *See* Jerry Kang, *Implicit Bias and Pushback from the Left*, 54 St. Louis L. Rev. 1139, 1146–48 (2010) (specifically rejecting complaint that implicit bias analysis must engage in reductionism).

27.   The number of stages is somewhat arbitrary. We could have listed more stages in a finer-grained timeline, or vice versa.

sistent with those implicit stereotypes. In other words, biases could shape whether an officer decides to stop an individual for questioning in the first place, elects to interrogate briefly or at length, decides to frisk the individual, and concludes the encounter with an arrest versus a warning.[28] These biases could contribute to the substantial racial disparities that have been widely documented in policing.[29]

Since the mid–twentieth century, social scientists have uncovered empirical evidence of negative attitudes toward African Americans as well as stereotypes about their being violent and criminal.[30] Those biases persist today, as measured by not only explicit but also implicit instruments.[31]

For example, Jennifer Eberhardt, Philip Goff, Valerie Purdie, and Paul Davies have demonstrated a bidirectional activation between Blackness and criminality.[32] When participants are subliminally primed[33] with a Black male face (as opposed to a White male face, or no prime at all), they are quicker to visually distinguish from background noise the faint outlines of a weapon.[34] In other words, by implicitly thinking "Black," they more quickly saw a weapon.

Interestingly, the phenomenon also happens in reverse. When subliminally primed with drawings of weapons, participants visually attended

---

28.   Devon W. Carbado, *(E)racing the Fourth Amendment*, 100 Mich. L. Rev. 946, 976-77 (2002).

29.   *See, e.g.,* Sam Vincent Meddis & Mike Snider, *Drug War 'Focused' on Blacks,* USA Today, Dec. 20, 1990, at 1A (reporting findings from a 1989 *USA Today* study that 41 percent of those arrested on drug charges were African American whereas 15 percent of the drug-using population is African American); Dianna Hunt, *Ticket to Trouble/Wheels of Injustice/Certain Areas are Ticket Traps for Minorities,* Houston Chron., May 14, 1995, at A1 (analyzing sixteen million Texas driving records and finding that minority drivers straying into White neighborhoods in Texas's major urban areas were twice as likely to get traffic violations than were Whites); Billy Porterfield, *Data Raise Question: Is the Drug War Racist?,* Austin Amer. Statesman, Dec. 4, 1994, at A1 (citing study showing African Americans more than seven times more likely to be arrested on drug charges than Whites in Travis County in 1993).

30.   *See generally,* Patricia G. Devine & Andrew J. Elliot, *Are Racial Stereotypes Really Fading? The Princeton Trilogy Revisited*, 21 Personality & Soc. Psychol. Bull. 1139 (1995).

31.   In a seminal paper, Patricia Devine demonstrated that being subliminally primed with stereotypically "Black" words prompted participants to evaluate ambiguous behavior as more hostile. *See* Patricia G. Devine, *Stereotypes and Prejudice: Their Automatic and Controlled Components*, 56 J. Personality & Soc. Psychol. 5 (1989). The priming words included "Negroes, lazy, Blacks, blues, rhythm, Africa, stereotype, ghetto, welfare, basketball, unemployed, and plantation." *Id.* at 10. Those who received a heavy dose of priming (80 percent stereotypical words) interpreted a person's actions as more hostile than those who received a milder dose (20 percent). *Id.* at 11–12. *See also* John A. Bargh et al., *Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action*, 71 J. Personality & Soc. Psychol. 230, 238–39 (1996).

32.   *See* Jennifer Eberhardt, et al., *Seeing Black: Race, Crime, and Visual Processing*, 87 J. Personality & Soc. Psychol. 876 (2004).

33.   The photograph flashed for only thirty milliseconds. *Id.* at 879.

34.   *See id.* at 879-80. There was a 21 percent drop in perceptual threshold between White face primes and Black face primes. This was measured by counting the number of frames (out of a total of 41) that were required before the participant recognized the outlines of the weapon in both conditions. There was a 8.8 frame difference between the two conditions. *Id.* at 881.

to Black male faces more than comparable White male faces.[35] This result was found not only in a student population, which is often criticized for being unrepresentative of the real world, but also among police officers.[36] The research suggests both that "Blackness" triggers weapons and makes them easier to see, and, simultaneously, that the idea of "weapons" triggers visual attention to Blackness. How these findings translate into actual police work is, of course, still speculative. But at a minimum, they suggest the possibility that officers have an implicit association between Blackness and weapons that could affect both their hunches and their visual attention.

Even if this is the case, one might respond that extra visual attention by the police isn't too burdensome. But who among us enjoys driving with a police cruiser on our tail?[37] Moreover, the increased visual attention didn't promote accuracy; instead, it warped the officers' perceptual memory. The subliminal prime of weapons led police officers not only to look more at Black faces but also to remember them in a biased way, as having more stereotypically African American features. Thus, they "were more likely to falsely identify a face that was more stereotypically Black than the target when they were primed with crime than when they were not primed."[38]

We underscore a point that is so obvious that it's easy to miss. The primes in these studies were all flashed *subliminally*. Thus the behavioral differences in visually attending to Black faces and also remembering them more stereotypically were all triggered implicitly, without the participants' conscious awareness.

Shooter bias. The implicit association between Blackness and weapons has also been found through other instruments, including other priming tasks[39] as well as the IAT. One of the tests available on Project Implicit specifically examines the implicit stereotype between African Americans

---

35.   Visual attendance was measured via a dot-probe paradigm, which requires participants to indicate on which side of the screen a dot flashes. The idea is that if you are already looking at one face (for example, the Black photograph), you will see a dot flash near the Black photograph faster. *See id.* at 881 (describing "dot-paradigm" as the gold standard in visual attention measures).

36.   *See id.* at 885–87 (describing methods, procedures and results of Study 4, which involved sixty-one police officers, 76 percent White, 86 percent male, and average age of forty-two). In this study, the crime primes were not pictures, but words: "violent, crime, stop, investigate, arrest, report, shoot, capture, chase, and apprehend." *Id.* at 886.

37.   *See* Carbado, *supra* note 26, at 966–67 (describing existential burdens of heightened police surveillance).

38.   Eberhardt et al., *supra* note 32, at 887.

39.   *See* B. Keith Payne, *Prejudice and Perception: The Role of Automatic and Controlled Processes in Misperceiving a Weapon*, 81 J. Personality & Soc. Psychol. 181, 185–86 (2001). The study deployed a "priming" paradigm, in which a photograph of a Black or White face was flashed to participants for two hundred milliseconds. Immediately thereafter, participants were shown pictures of guns or tools. *Id.* at 184. By being primed by the Black face, participants were faster to identify guns. *Id.* at 185.

(as compared to European Americans) and weapons (as compared to harmless items). That association has been found to be strong, widespread, and also dissociated from explicit self-reports.[40]

Skeptics can reasonably ask why we should care about small time differentials between schema-consistent and -inconsistent pairings that are often no more than half a second. But it's worth remembering that half a second may be all the time a police officer has to decide whether to shoot. In the policing context, that half a second might mean the difference between life and death.

Joshua Correll developed a "shooter paradigm," in which participants are confronted with photographs of individuals (targets) holding an object, superimposed on various city landscapes.[41] On the one hand, if the object is a weapon, the participant is instructed to press a key to "shoot." On the other hand, if the object is harmless (for example, a wallet), the participant must press a different key to "holster" the weapon. Correll found that participants were quicker to press "shoot" when the target was Black as compared to White.[42] Also, under time pressure, participants made more mistakes (false alarms) and shot more unarmed Black targets than unarmed White targets, and failed to shoot more armed White targets (misses) than armed Black targets.[43] Interestingly, the shooter bias effect was not correlated with measures of explicit personal stereotypes.[44] Correll also found comparable amounts of shooter bias in African American participants.[45] This suggests that negative attitude toward African Americans isn't what's driving the phenomenon.[46]

---

40.   For N=85,742 participants, the average IAT D score was 0.37. If the IAT D score is divided by its standard deviation of the population that has taken the test, the result is interpretable as the commonly used effect size measure, Cohen's *d*, which was 1.00. By contrast, the self-reported association (that is, explicit stereotype measure) was Cohen's *d*=0.31. *See* Nosek et al., *supra* note 13, at 11, tbl.2.

41.   Joshua Correll et al., *The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals*, 83 J. Personality. & Soc. Psychol. 1314, 1315–17 (2002) (describing procedure).

42.   *Id.* at 1317.

43.   *Id.* at 1319. For qualifications about how the researchers discarded outliers, see Jerry Kang, *Trojan Horses of Race*, 118 Harv. L. Rev. 1489, n.16 (2005). Correll's general findings have been confirmed in subsequent studies. *See, e.g.*, Anthony G. Greenwald et al., *Targets of Discrimination: Effects of Race on Responses to Weapons Holders*, 39 J. Experimental Soc. Psychol. 399 (finding similar results).

44.   Correll, *supra* note 39, at 1323. The scales used were the Modern Racism Scale, the Discrimination and Diversity Scale, the Motivation to Control Prejudiced Responding Scale, and some questions from the Right-Wing Authoritarianism Scale and the Personal Need for Structure Scale for good measure. *Id.* at 1321. Shooter bias was, however, correlated with measures of societal stereotypes—the stereotypes that other people supposedly held. *Id.* at 1323.

45.   *See id.* at 1324.

46.   On explicit attitude instruments, African Americans show on average substantial in-group preference (over Whites). On implicit attitude instruments, such as the IAT, African Americans show on average no preference, which means that they bell curve around zero. *See* Brian A. Nosek, Mahzarin R. Banaji & Anthony G. Greenwald, *Harvesting Implicit Group Attitudes and Beliefs From a Demonstration Web Site*, 6 Group Dynamics 101, l05–06 (2002).

The shooter bias experiments have also been run on actual police officers, with some similar and some different results. In one study, police officers showed the same bias in favor of shooting unarmed Blacks more often than unarmed Whites that student and civilian populations demonstrated.[47] However, in another study, although police officers showed a similar speed bias, they did not show any racial bias in the most important criterion of accuracy. In other words, there was no higher error rate of shooting unarmed Blacks as compared to Whites.[48]

Finally, in a study that directly linked implicit stereotypes (with weapons) as measured by the IAT and "shooter bias," Jack Glaser and Eric Knowles found that "[i]ndividuals possessing a relatively strong stereotype linking Blacks and weapons [1 standard deviation above the mean IAT] clearly show the Shooter Bias . . . ."[49] By contrast, recall that Correll found no such correlation with explicit stereotypes. These findings are consistent with the implicit stereotype story. Of course, it may also be true that participants were simply downplaying or concealing their explicit bias, which could help explain why no correlation was found.

In sum, we have evidence that suggests that implicit biases could well influence various aspects of policing. A fairly broad set of research findings shows that implicit biases (as measured by implicit instruments) alter and affect numerous behaviors that police regularly engage in—visual surveillance, recall, and even armed response.[50] The general social psychological insight that emerges is the implicit form of what legal academics now regularly call "confirmation bias": that police will interpret ambiguous information in accordance with their (implicit as well as explicit) stereotypes of the populations they encounter.

It should go without saying that explicit biases, which often undergird unspoken policies of racial profiling of entire neighborhoods and racial minorities, also play an enormous role in differential policing of people of color. It also should go without saying that various structural forces that produce racially segregated, predominantly minority neighborhoods that are higher in poverty and crime also have a huge impact on racialized policing. Nevertheless, we repeat these points so that readers internalize that implicit, explicit, and structural processes should not be deemed mutually exclusive.

---

47.   *See* E. Ashby Plant & B. Michelle Peruche, *The Consequences of Race for Police Officers' Responses to Criminal Subjects,* 16 Psychol. Sci. 180, 181 (2005).

48.   *See* Joshua Correll et al. *Across the Thin Blue Line: Police Officers and Racial Bias in the Decision to Shoot,* 92 J. Personality & Soc. Psychol. 1006, 1010–13, 1016–17 (2007) (describing results from two studies).

49.   Jack Glaser & Eric D. Knowles, *Implicit Motivation to Control Prejudice*, 44 J. Experimental Soc. Psychol. 164 , 169 (2008).

50.   For discussions in the law reviews, with some treatment of implicit biases, see Alex Geisinger, *Rethinking Profiling: A Cognitive Model of Bias and Its Legal Implications*, 86 Or. L. Rev. 657, 667–73 (2007) (providing cognitive model based on automatic categorization, in accordance with "behavioral realism").

2.   Charge and Plea Bargain

Journalistic investigations have uncovered some statistical evidence that racial minorities are treated worse than Whites in the charging decisions of prosecutors.[51] Of course, there might be some legitimate reason for those disparities, for example, if minorities and Whites are not similarly situated on average.[52] One way to examine whether the merits drive the disparate results is to control for everything except some irrelevant attribute, such as race. In several studies, researchers used regression analyses to conclude that race was indeed independently correlated with the severity of the prosecutor's charge.

For example, in a 1985 study of charging decisions by prosecutors in Los Angeles, researchers found prosecutors more likely to press charges against Black than White defendants, and determined that these charging disparities could not be accounted for by race neutral factors such as prior record, seriousness of charge, or use of a weapon.[53] Two studies, one in Florida and the other in Indiana, also in the late 1980s found charging discrepancies based upon the race of the victim.[54] At the federal level, a U.S. Sentencing Commission report found that prosecutors were more apt to offer White defendants generous plea bargains with sentences below the prescribed guidelines than Black or Latino defendants.[55]

---

51.   For example, in San Jose, a newspaper investigation concluded that out of the almost seven hundred thousand criminal cases reported, "at virtually every stage of pre-trial negotiation, whites are more successful than non-whites." Ruth Marcus, *Racial Bias Widely Seen in Criminal Justice System; Research Often Supports Black Perceptions,* Wash. Post May 12, 1992, at A4. San Francisco Public Defender Jeff Brown commented on racial stereotyping: "It's a feeling, 'You've got a nice person screwing up,' as opposed to feeling that 'this minority is on a track and eventually they're going to end up in state prison.'" Christopher H. Schmitt, *Why Plea Bargains Reflect Bias,* San Jose Mercury News, Dec. 9, 1991, at 1A; Christopher Johns, *The Color of Justice: More and More, Research Shows Minorities Aren't Treated the Same as Anglos by the Criminal Justice System,* Ariz. Republic, July 4, 1993, at C1 (citing several reports showing disparate treatment of blacks in the criminal justice system).

52.   This, of course, might be a product of structural bias, but many consider such bias much less troubling than explicit or implicit biases, which cause different treatment of otherwise identical individuals simply because of the social categories they belong to.

53.   *See* Michael L. Radelet & Glenn L. Pierce, *Race and Prosecutorial Discretion in Homicide Cases*, 19 L. & Soc'y Rev. 587, 615–19 (1985).

54.   *See* Kenneth B. Nunn, *Ethics in Criminal Advocacy, Symposium, The "Darden Dilemma": should African Americans Prosecure Crimes?*, 68 Fordham L. Rev. 1473, 1493 (citing Michael L. Radelete & Glenn L. Pierce, *supra* note 51, at 615-19; Martha A. Myers & John Hagan, *Private and Public Trouble: Prosecutors and the Allocation of Court Resources*, 26 Soc. Probs. 439, 441-47 (1979)).

55.   *Justice on Trial, Chapter Two: Race and Prosecutorial Discretion,* The Leadership Conference, http://www.civilrights.org/publications/justice-on-trial/prosecutorial.html (citing United States Sentencing Comm'n, Special Report to Congress: Cocaine and Federal Sentencing Policies); *see also* Kevin McNally, *Race and Federal Death Penalty: A Nonexistent Problem Gets Worse*, 53 DePaul L. Rev. 1615 (2004) (compiling studies on the death penalty).

While these studies are suggestive, other studies find no disparate treatment.[56] Moreover, this kind of statistical evidence doesn't definitively tell us that biases generally or implicit biases specifically produce discriminatory charging decisions or plea offers by prosecutors, or a discriminatory willingness to accept worse plea bargains on the part of defense attorneys. The best way to get evidence on such hypotheses would be to measure the implicit bias of prosecutors and defense attorneys and investigate the extent to which that bias predicts different treatment of cases otherwise identical on the merits.

Unfortunately, we have very little data on this front. Indeed, we have no studies, as of yet, that look at prosecutors' and defense attorneys' implicit biases and attempt to correlate them with those individuals' charging practices or plea bargains. Nor do we know as much as we'd like about their implicit biases more generally. But on that score, we do know something. Let's start with defense attorneys. One might think that defense attorneys, repeatedly put into the role of interacting with what is often a disproportionately minority clientele, and often ideologically committed to racial equality,[57] might have materially different implicit biases from the general population. But Ted Eisenberg and Sheri Lynn Johnson found evidence to the contrary: Even capital punishment defense attorneys show negative implicit attitudes toward African Americans.[58] Their implicit attitudes toward Blacks roughly mirrored those of the population at large.

What about prosecutors? To our knowledge, no one has measured specifically the implicit biases held by prosecutors.[59] That said, there's no

---

56. *See, e.g.,* Cyndy Caravelis, Ted Chiricos & William Bales, *Race, Ethnicity, Threat, and the Designation of Career Offenders*, 2011 Just. Q. 1 (2011) (showing that in some counties, Blacks and Latinos are more likely to be prosecuted as career offenders, but in other counties with different demographics, Blacks and Latinos have a lesser likelihood of being prosecuted as career offenders than do Whites with similar profiles); Jeremy D. Ball, *Is It a Prosecutor's World? Determinants of Count Bargaining Decisions*, 22 J. Contemp. Crim. Just. 241 (2006) (finding no correlation between race and the willingness of prosecutors to reduce charges in order to obtain guilty pleas, but acknowledging that study did not include evaluation of original arrest report).

57. *See* Gordon B. Moskowitz, Amanda R. Salomon & Constance M. Taylor, *Preconsciously Controlling Stereotyping: Implicitly Activated Egalitarian Goals Prevent the Activation of Stereotypes*, 18 Soc. Cognition 151, 155-56 (2000) (showing that "chronic egalitarians" who are personally committed to removing bias in themselves do not exhibit implicit attitudinal preference for Whites over Blacks).

58. *See* Theodore Eisenberg & Sheri Lynn Johnson, *Implicit Racial Attitudes of Death Penalty Lawyers*, 53 DePaul L. Rev. 1539, 1545-55 (2004). The researchers used a Black/White attitude paper-pencil IAT. *Id.* at 1543-45. The defense attorneys displayed biases that were comparable to the rest of the population. *Id.* at 1553.

59. In some contexts, prosecutors have resisted revealing information related potentially to their biases. For example, in *United States v. Armstrong*, 517 U.S. 456 (1996), defendants filed a motion to dismiss the indictment for selective prosecution, arguing that the U.S. Attorney prosecuted virtually all African Americans charged with crack offenses in federal court, but left all White crack defendants to be prosecuted in state court, resulting in much longer sentences for identical offenses. *Id.* at 460-61. The claim foundered when the U.S. Attorney's Office resisted the defendants' discovery motion concerning criteria for prosecutorial decisions and the U.S. Supreme Court upheld the U.S. Attorney's Office's refusal to provide discovery. *Id.* at 459-62. The Court

reason to presume attorney exceptionalism in terms of implicit biases.[60] And if defense attorneys, who might be expected to be less biased than the population, show typical amounts of implicit bias, it would seem odd to presume that prosecutors would somehow be immune. If this is right, there is plenty of reason to be concerned about how these biases might play out in practice.

As we explain in greater detail below, the conditions under which implicit biases translate most readily into discriminatory behavior are when people have wide discretion, exercised quickly, with little accountability. Prosecutors function in just such environments.[61] They exercise tremendous discretion, on whom, whether, and at what level of severity to charge a particular crime, as well as influence the terms of and likelihood of a plea bargain and the length of the prison sentence—all with little judicial oversight. Other psychological theories—such as confirmation bias, social judgeability theory, and shifting standards, which we discuss below[62]—reinforce our hypothesis that prosecutorial decisionmaking indeed risks being influenced by implicit bias.

3.   Trial

a.   Jury

If the case goes to the jury, what do we know about how implicit biases might influence the factfinder's decisionmaking? There is a long line of research on racial discrimination by jurors, mostly in the criminal context. Notwithstanding some mixed findings, the general research consensus is that jurors of one race tend to show bias against defendants who belong to another race ("racial outgroups"). For example, White jurors will treat Black defendants worse than they treat comparable White defendants. The best and most recent meta-analysis of laboratory juror studies was performed by Tara Mitchell and colleagues, who found that the fact that a juror was of a different race than the defendant influenced both verdicts

---

held that, prior to being entitled even to discovery, defendants claiming selective prosecution cases based on race must produce credible evidence that "similarly situated individuals of a different race were not prosecuted." *Id.* at 465.

    60.   Several of the authors have conducted training sessions with attorneys in which we run the IAT in the days leading up to the training. The results of these IATs have shown that attorneys harbor similar biases to the rest of the population. One recent study of a related population, law students, confirmed that they too harbor implicit gender biases. *See* Justin D. Levinson & Danielle Young, *Implicit Gender Bias in the Legal Profession: An Empirical Study*, 18 Duke J. Gender L. & Pol'y 1, 28–31 (2010).

    61. *See* Robert J. Smith & Justin D. Levinson, *The Impact of Implicit Racial Bias on the Exercise of Prosecutorial Discretion*, Seattle L. Rev. (forthcoming, 2012)(on file with authors) (undertaking a step-by-step consideration of how prosecutorial discretion may be fraught with implicit bias).

    62.   *See infra* Part II.B.

and sentencing.[63] The magnitude of the effect sizes were measured conservatively[64] and found to be small (Cohen's d=0.092 for verdicts, d=0.185 for sentencing).[65]

But effects deemed "small" by social scientists may nonetheless have huge consequences for the individual, the social category he belongs to, and the entire society. For example, if White juries rendered guilty verdicts in exactly 80 percent of their decisions,[66] then an effect size of Cohen's d=0.095 would mean that the rate of conviction of Black defendants will be 83.8 percent, compared to 76.2 percent for White defendants. Put another way, in one hundred otherwise identical trials, eight more Black than White defendants would be found guilty.[67]

One might assume that juror bias against racial outgroups would be greater when the case is somehow racially charged or inflamed, as opposed to those instances when race does not explicitly figure in the crime. Interestingly, many experiments have demonstrated just the opposite.[68] Sam Sommers and Phoebe Ellsworth explain the counterintuitive phenomenon in this way: When the case is "racially charged," jurors—who want to be fair—respond by being more careful and thoughtful about race and their own assumptions and thus do not show bias in their deliberations and outcomes. By contrast, when the case is not racially charged, even though there is a Black defendant and a White victim, jurors are not especially vigilant about the possibility of racial bias influencing their decisionmaking. These findings are more consistent with an implicit bias story than a concealed explicit bias story.[69]

---

63.   Tara L. Mitchell, et al., *Racial Bias in Mock Juror Decision-Making: A Meta-Analytic Review of Defendant Treatment*, 29 L. & Hum. Behav. 621, 627–28 (2005). The meta-analysis processed thirty-four juror verdict studies (with 7397 participants) and sixteen juror sentencing studies (with 3141 participants). *Id.* at 625. All studies involved experimental manipulation of the defendant's race. Multirace participant samples were separated out in order to maintain the study's definition of racial bias as differential treatment by a juror of a defendant who belonged to a racial outgroup. *See Id.*

64.   Studies that reported nonsignificant results (p > .05) for which effect sizes could not be calculated were given effect sizes of .00. *Id.*

65.   *Id.* at 629.

66.   *See* Tracy Kyckelhahn & Thomas H. Cohen, *Felony Defendants in Large Urban Counties*, *in* Bureau of Just. Stat. Bull. 1, 3 (April 2008), *available at* http://bjs.ojp.usdoj.gov/content/pub/pdf/fdluc04.pdf ("Seventy-nine percent of trials resulted in a guilty verdict or judgment, including 82% of bench trials and 76% of jury trials."). *See also* Thomas H. Cohen & Tracey Kyckelhahn, *Felony Defendants in Large Urban Counties, 2006*, in Bureau of Just. Stat. Bull. 1 (May 2010) *available at* http://bjs.ojp.usdoj.gov/content/pub/pdf/ fdluc06.pdf (reporting the "typical" outcome as three out of four trials resulting in convictions).

67.   This translation between effect size *d* values and outcomes was described by Robert Rosenthal & Donald B. Rubin, *A Simple, General Purpose Display of Magnitude of Experimental Effect*, 74 J. Educ. Psychol. 166 (1982).

68.   *See, e.g.*, Samuel R. Sommers & Phoebe C. Ellsworth, *"Race Salience" in Juror Decision-Making: Misconceptions, Clarifications, and Unanswered Questions*, 27 Behav. Sci. & L. 599 (2009).

69.   *See* Samuel R. Sommers & Phoebe C. Ellsworth, *White Juror Bias: An Investigation of Prejudice Against Black Defendants in the American Courtroom*, 7 Psychol. Pub. Pol'y & L. 201, 255

So far, we know that race effects have been demonstrated in juror studies (sometimes in counterintuitive ways), but admittedly little is known about "the precise psychological processes through which the influence of race occurs in the legal context."[70] Our default assumption is juror unexceptionalism—given that implicit biases generally influence decisionmaking, there's no reason to presume that citizens become immune to the effects of these biases when they serve in the role of jurors. Leading scholars from the juror bias field have expressly raised the possibility that the psychological mechanisms might be "unintentional and even non-conscious processes."[71]

Some recent juror studies by Justin Levinson and Danielle Young have tried to disentangle the psychological mechanisms of juror bias by using the IAT and other methods. In one mock juror study, Levinson and Young had participants view five photographs of a crime scene, including a surveillance camera photo that featured a masked gunman whose hand and forearm were visible. For half the participants, that arm was dark-skinned; for the other half, that arm was lighter-skinned.[72] The participants were then provided twenty different pieces of trial evidence. The evidence was designed to produce an ambiguous case regarding whether the defendant was indeed the culprit. Participants were asked to rate how much the presented evidence tended to indicate the defendant's guilt or innocence, and to decide whether the defendant was guilty or not, using both a guilty/not guilty scale as well as a 0 to 100 likelihood scale.

The study found that the subtle manipulation of the skin color altered how jurors evaluated the evidence presented and also how they answered the crucial question "How guilty is the defendant?" (dark skin guilt mean score, M=66.97; light skin guilt, M=56.37, with 100 being "definitely guilty").[73] Measures of explicit bias, including the Modern Racism Scale and feeling thermometers, showed no statistically significant correlation with the participants' weighing of the evidence or assessment of guilt. More revealing, participants were asked to recall the race of the masked robber (which was a proxy for the light or dark skin), but many could not

---

(2001); Samuel R. Sommers & Phoebe C. Ellsworth, *Race in the Courtroom: Perceptions of Guilt and Dispositional Attributions*, 26 Personality & Soc. Psychol. Bull. 1367–79 (2000). That said, one could still hold to an explicit bias story in the following way. The juror has a negative attitude and/or stereotype that he is consciously aware of and endorses. But he knows it's not socially acceptable. So he conceals it. When a case is "racially charged," racial bias is more salient, so other jurors will be on the lookout for bias. Accordingly, the juror conceals it even more, all the way up to making sure that his behavior is completely race-neutral. This explicit bias story is not mutually exclusive with the implicit bias story we are telling.

70.   Samuel R. Sommers, *Race and the Decision-Making of Juries*, 12 Legal & Criminological Psych. 171, 172 (2007).

71.   *Id.* at 175.

72.   Levinson & Young, *supra* note 18, at 332–33 (describing experimental procedures).

73.   *See id.* at 337 (confirming that the difference was statistically significant, F=4.40; p=0.034, *d=0.52*).

recall.[74]   Moreover, their recollections didn't correlate with their judgments of guilt. Taken together, these findings suggest that implicit bias—not explicit, concealed bias, or even any degree of conscious focus on race—was influencing how jurors assessed the evidence in the case.

In fact, there's even clearer evidence that implicit bias was at work. Levinson, Huajian Cai, and Young also constructed a new IAT, the "Guilty/ Not Guilty" IAT, to test implicit stereotypes of African Americans as guilty (not innocent).[75] They gave the participants this new IAT, as well as the general race attitude IAT. They found that participants showed an implicit negative attitude toward Blacks as well as a small implicit stereotype between Black and Guilty.[76] More important than the bias itself is whether it predicts judgment. On the one hand, regression analysis demonstrated that a measure of *evidence evaluation* was a function of both the implicit attitude and the implicit stereotype.[77] On the other hand, the IAT scores did not predict what's arguably more important: guilty verdicts or guilty scale judgments.[78] In sum, a subtle change in skin color changed judgments of evidence and guilt; implicit biases measured by the IAT predicted how identical pieces of information were evaluated.

We have a long line of juror research, as synthesized through a meta-analysis revealing that jurors of one race treat defendants of another race worse on verdict and sentencing. According to some experiments, that difference might take place *more* often in experimental settings when the case is *not* racially charged, which suggests that participants who seek to be fair will endeavor to correct for potential bias when the threat of potential race bias is obvious. Finally, some recent work reveals that certain IATs can predict racial discrimination in the evaluation of evidence by mock jurors. Unfortunately, because of the incredible difficulties in research design, we do not have studies looking at implicit bias in real criminal trials. Accordingly, the existing body of research, while strongly suggestive, pro-

---

74.   This finding built upon Levinson's previous experimental study of implicit memory bias in legal decisionmaking. *See* Justin D. Levinson, *Forgotten Racial Equality: Implicit Bias, Decisionmaking, and Misremembering*, 57 DUKE L.J. 345, 398–406 (2007) (finding that study participants misremembered trial relevant facts in racially biased ways).

75.   Justin D. Levinson, Huajian Cai & Danielle Young, *Guilty by Implicit Bias: The Guilty–Not Guilty Implicit Association Test*, 8 OHIO ST. J. CRIM. L. 187 (2010).

76.   *Id.* at 204. For the attitude IAT, D=0.21 (*p*<0.01). Id. at 204 n.87. For the Guilty–Not Guilty IAT, D=0.18 (p<0.01). *Id.* at 204 n.83.

77.   Participants rated each of the twenty pieces of information (evidence) in terms of its probity regarding guilt or innocence on a 1–7 scale. This produced a total "evidence evaluation" score that could range between 20 (least amount of evidence of guilt) to 140 (greatest). *Id.* at 202 n.70 (citation omitted). The greater the "Black=guilty" stereotype or the greater the negative attitude toward Blacks, the higher the guilty "evidence evaluation." The ultimate regression equation was: Evidence = 88.58 + 5.74 x BW + 6.61 x GI + 9.11 x AI + e (where BW stands for Black or White suspect; GI stands for guilty stereotype IAT score; AI stands for race attitude IAT score; *e* stands for error). *Id.* at 206. In normalized units, the implicit stereotype β=0.25 (p<0.05); the implicit attitude β=0.34 (p<0.01); adjusted R^2=0.24. *See id.* nn. 93–95.

78.   *Id.* at n.95.

vides inferential rather than direct support that implicit bias accounts for some of the race effects on conviction and sentencing.

b.   Judge

Obviously, the judge plays a crucial role in various aspects of the trial, exercising important discretion in setting bail,[79] deciding motions, conducting and deciding what can be asked during jury selection, ruling on the admissibility of evidence, presiding over the trial, and rendering verdicts in some cases. Again, as with the lawyers, there is no inherent reason to think that judges are immune from implicit biases. The extant empirical evidence supports this assumption as well.[80] Jeff Rachlinski and his coauthors are the only researchers who have measured the implicit biases of actual trial court judges. They have given the race attitude IAT to judges from three different judicial districts. Consistent with the general population, the White judges showed strong implicit attitudes in favor of Whites over Blacks.[81]

Rachlinski and colleagues tried to see if these biases predicted behavioral differences by giving judges three different vignettes and asking for their views on various questions, ranging from the likelihood of defendant recidivism to recommended verdict and confidence level. Two of these vignettes said nothing about race although some of the judges were subliminally primed with words designed to trigger the social category "African American." The third vignette explicitly identified the defendant (and victim) as White or Black and did not use subliminal primes. After collecting the responses, Rachlinski et al. analyzed whether judges treated White or Black defendants differently and whether any such difference could be predicted by the IAT.

They found mixed results. In the two subliminal priming vignettes, judges didn't respond differently as a function of the primes on average. In other words, the primes didn't prompt them to be harsher on defendants across the board, as prior priming studies with nonjudge populations

---

79.   *See* Ian Ayres & Joel Waldfogel, *A Market Test for Race Discrimination in Bail Setting*, 46 Stan. L. Rev. 987, 992 (1994) (finding 35 percent higher bail amounts for Black defendants after controlling for eleven other variables besides race).

80. Judge Bennett, a former civil rights lawyer, shares his unnerving discovery of his own disappointing IAT results in Mark W. Bennett, *Unraveling the Gordian Knot of Implicit Bias in Jury Selection: The Problems of Judge-Dominated Voir Dire, the Failed Promise of* Batson*, and Proposed Solutions*, 4 Harv. L. & Pol'y Rev. 149 ,150 (2010).

81.   *See* Jeffrey J. Rachlinski, et al., *Does Unconscious Racial Bias Affect Trial Judges?*, 84 Notre Dame L. Rev. 1195, 1210 (2009). White judges (N=85) showed an IAT effect M=216 ms (with a standard deviation of 201 ms). 87.1 percent of them were quicker to sort in the schema-consistent arrangement versus the schema-inconsistent one. Black judges (N=43) showed a small bias M=26 ms (with a standard deviation of 208 ms). Only 44.2 percent of Black judges were quicker to sort in the schema-consistent arrangement versus the schema-inconsistent one. *See id.*

had found.[82] That said, the researchers found a *marginally* statistically sig-
nificant interaction with IAT scores: Judges who had a greater degree of
implicit bias against Blacks (and relative preference for Whites) were
harsher on defendants (who were never racially identified) when they had
been primed (with the Black words). By contrast, those judges who had
implicit attitudes in favor of Blacks were less harsh on defendants when
they received the prime.[83]

In the third vignette, a battery case that explicitly identified the de-
fendant as one race and the victim as the other,[84] the White judges showed
equal likelihood of convicting the defendant, whether identified as White
or Black. By contrast, Black judges were much more likely to convict the
defendant if he was identified as White as compared to Black. When the
researchers probed more deeply to see what, if anything, the IAT could
predict, they didn't find the sort of interaction that they found in the other
two vignettes—in other words, judges with strong implicit bias in favor of
Whites didn't treat the Black defendant more harshly.[85]

Noticing the difference between White and Black judge responses in
the third vignette study, the researchers probed still deeper and found a
three-way interaction between judge's race, judge's IAT score, and de-
fendant's race. No effect was found for White judges; the core finding
concerned, instead, Black judges. Those Black judges with a stronger
Black preference on the IAT were less likely to convict the Black defend-
ant (as compared to the White defendant); correlatively, those Black judg-
es with White preference on the IAT were more likely to convict the Black
defendant.[86]

It's hard to make simple sense of such complex findings, which may
have been caused in part by the fact that the judges quickly sniffed out the
purpose of the study—to detect racial discrimination.[87] Given the high
motivation not to perform race discrimination under research scrutiny, one
could imagine that White judges might make sure to correct for any poten-
tial unfairness. By contrast, Black judges may have felt less need to signal
racial fairness, which might explain why Black judges showed different be-
haviors as a function of implicit bias whereas White judges did not.

Put another way, data show that when the race of the defendant is ex-
plicitly identified to judges in the context of a psychology study (that is,
the third vignette), judges are strongly motivated to be fair, which prompts

---

82.   *See* Sandra Graham & Brian S. Lowery, *Priming Unconscious Racial Stereotypes About Adolescent Offenders*, 28 L. & Hum. Behav. 483 (2004).
83.   *See* Rachlinski et al., *supra* note 81, at 1215. An ordered logit regression was performed between the judge's disposition against the priming condition, IAT score, and their interaction. The interaction term was marginally significant at *p*=0.07. *See id.* at n.94.
84.   This study did not use any subliminal primes.
85.   *See id.* at n.41.
86.   *Id.* at 1220 n. 114.
87.   *See id.* at 1223.

a different response from White judges (who may think to themselves "whatever else, make sure not to treat the Black defendants worse") than Black judges (who may think "give the benefit of the doubt to Black defendants"). However, when race is not explicitly identified but implicitly primed (vignettes one and two), perhaps the judges' motivation to be accurate and fair is not on full alert. Notwithstanding all the complexity, this study provides some suggestive evidence that implicit attitudes may be influencing judges' behavior.

4. Sentencing

There is evidence that African Americans are treated worse than similarly situated Whites in sentencing. For example, federal Black defendants were sentenced to 12 percent longer sentences under the Sentencing Reform Act of 1984,[88] and Black defendants are subject disproportionately to the death penalty.[89] Of course, it's possible that there's some good reason for that difference, based on the merits. One way to check is to run experimental studies holding everything constant except for race.

Probation officer. In one study, Sandra Graham and Brian Lowery subliminally primed police officers and juvenile probation officers with words related to African Americans, such as "Harlem" or "dreadlocks." This subliminal priming led the officers to recommend harsher sentencing decisions.[90] As we noted above, Rachlinski et al. found no such effect on the judges they tested using a similar but not identical method.[91] But at least

---

88. *See* David B. Mustard, *Racial, Ethnic, and Gender Disparities in Sentencing: Evidence From the U.S. Federal Courts*, 44 J.L. & ECON. 285, 300 (2001) (examining federal judge sentencing under the Sentencing Reform Act of 1984).

89. U.S. GEN. ACCOUNTING OFFICE., REPORT TO THE SENATE AND HOUSE COMMITTEES ON THE JUDICIARY, DEATH PENALTY SENTENCING: RESEARCH INDICATES PATTERN OF RACIAL DISPARITIES (Feb. 1990) (finding killers of White victims receive the death penalty more often than killers of Black victims); David Baldus et al., *Racial Discrimination and the Death Penalty in the Post–Furman Era: An Empirical and Legal Overview, with Recent Findings from Philadelphia*, 83 CORNELL L. REV. 1638, 1710-24 (1998) (finding mixed evidence that Black defendants are more likely to receive the death sentence).

90. *See* Sandra Graham & Brian Lowery, *Priming Unconscious Racial Stereotypes about Adolescent Offenders*, 28 L. & HUM. BEHAV. 483 (2004).

91. Priming studies are quite sensitive to details. For example, the more subliminal a prime is (in time duration, frequency), the less the prime tends to "stick" (the smaller the effects, the faster it dissipates). Rachlinski et al. identified some differences between their experimental procedure and that of Graham and Lowery's. *See* Rachlinski et al., *supra* note 81, at 1213 n.88. Interestingly, in the Rachlinski study, for judges from the "eastern conference" (seventy judges), a programming error made their subliminal primes last only sixty-four milliseconds. By contrast, for the "western conference" (forty-five judges) the prime lasted 153 milliseconds, which was close to the duration used by Graham and Lowery (150 milliseconds). *See id.* at 1206 (providing numerical count of judges); *id.* at 1213 n.84 (identifying programming error). Graham and Lowery wrote that they selected the priming durations through extensive pilot testing "to arrive at a presentation time that would allow the primes to be detectable but not identifiable." Graham & Lowery, *supra* note 90, at 489. It's possible that the truncated priming duration for the "eastern conference" judges contributed to the different findings between Rachlinski et al. and Graham and Lowery.

in this study, an effect was found with police and probation officers. Given that this was a subliminal prime, the merits could not have justified the different evaluations.

Afrocentric features. Irene Blair, Charles Judd, and Kristine Chapleau took photographs from a database of convicted criminals from the State of Florida[92] and asked participants to judge how "Afrocentric" both White and Black inmates looked on a one to nine scale.[93] The goal was to see if race and/or facial features correlated with actual sentencing. Using multiple regression analysis, the researchers found that after controlling for the seriousness of the primary and additional offenses, the race of the defendant showed no statistical significance. In other words, White and Black defendants were sentenced without discrimination based on race. According to the researchers, this is a success story based on various sentencing reforms specifically adopted by the State of Florida mostly to decrease sentencing discretion.

However, when the researchers added Afrocentricity of facial features into their regressions, they found a curious correlation. Within each race, either Black or White, the more Afrocentric the defendant looked, the harsher his punishment.[94] How much so? If you picked a defendant who was one standard deviation above the mean in Afrocentric features and compared him to another defendant of the same race who was one standard deviation below the mean, there would be a sentence difference of seven to eight months between them, holding constant any difference in their actual crime.[95]

Again, if the research provides complex findings, we must grapple with a complex story. On the one hand, we have good news: Black and White defendants were, overall, sentenced comparably. On the other hand, we have bad news: Within each race, the more "Black" you look, the harsher your punishment. What might make sense of such results? According to the researchers, perhaps implicit bias was responsible. If judges are motivated to show no racial discrimination, they may be on guard regarding the dangers of treating similarly situated Blacks worse

---

92. *See* Irene V. Blair et al., *The Influence of Afrocentric Facial Features in Criminal Sentencing*, PSYCHOL. SCI. 614, 675 (2004) (selecting a sample of 100 Black inmates; 116 White inmates).

93. *Id.* at 676. Afrocentric means full lips, broad nose, relatively darker skin color, and curly hair. It is what participants socially "understand" to look African without any explicit instruction or definition. *See id.* at 674 n.1.

94. *Id.* at 676–77. Jennifer Eberhardt and her colleagues reached consistent findings when she used the same Florida photograph dataset to examine how Black defendants were sentenced to death. After performing a median split on how stereotypical the defendant looked, the top half were sentenced to death 57.5 percent of the time compared to the bottom half, which were sentenced to death only 24.4 percent of the time. *See* Jennifer L. Eberhardt et al., *Looking Deathworthy: Perceived Stereotypicality of Black Defendants Predicts Capital-Sentencing Outcomes*, 17 PSYCHOL. SCI. 383, 384 (2006). Interestingly, this effect was not seen when the victim was Black. *See id.* at 385.

95. *See* Blair et al., *supra* note 92, at 677-78.

than Whites. On alert to this potential bias, the judges prevent it from causing any discriminatory behavior. By contrast, judges have no conscious awareness that Afrocentric features might be triggering stereotypes of criminality and violence that could influence their judgment. Without such awareness, they could not explicitly control or correct for the potential bias.[96] If this explanation is correct, we have further evidence that discrimination is being driven in part by implicit and not solely explicit-but-concealed biases.

* * *

Where does this whirlwind tour of psychological research findings leave us? In each of the stages of the criminal trial process discussed, the empirical research gives us reason to think that implicit biases—attitudes and beliefs that we are not directly aware of and may not endorse—could influence how defendants are treated and judged. Wherever possible, in our description of the studies, we have tried to provide the magnitude of these effects. But knowing precisely how much work they really do is difficult. If we seek an estimate, reflective of an entire body of research and not any single study, one answer comes from the Greenwald meta-analysis, which found that the IAT (which is the most widely used, but not the only measure of implicit bias) could predict 5.6 percent of the variation of the behavior in Black-White behavioral domains.[97]

Should that be deemed a lot or a little? In answering this question, we should be mindful of the collective impact of such biases, integrated over time (per person) and over persons (across all defendants).[98] For a single defendant, these biases may surface for various decisionmakers repeatedly in the stages of policing, charging, bail, plea bargaining, pretrial motions, evidentiary motions, witness credibility, lawyer persuasiveness, guilt determination, sentencing recommendations, sentencing itself, appeal, and so on. Even small biases at each stage may snowball into a substantial effect in aggregate.

To get a more concrete sense, Anthony Greenwald has produced a simulation that models cumulating racial disparities through five sequential stages of criminal justice—arrest, arraignment, plea bargain, trial, and sentence. It supposes that the probability of arrest having committed the offense is .50; that the probability of conviction at trial is .75; and that the effect size of implicit bias to be r=0.1 at each stage. Under this simulation,

---

96.   *See id.* at 677.

97.See Greenwald et al., *supra* note 21, at 24 tbl. 3 (showing that correlation between Race attitude IAT (Black/White) and behavior in the meta-analysis is .236, which when squared equals 0.056, which is the percentage of variance explained).

98.   *See* Rachlinski et al., *supra* note 81, at 1202; Jerry Kang & Mahzarin Banaji, *Fair Measures: A Behavioral Realist Revision of 'Affirmative Action'*, 94 Calif. L. Rev. 1063, 1073 (2006).

for a crime whose mean sentence is 5 years, with a standard deviation of 2 years, Black criminals can expect a sentence of 2.44 years versus White criminals who can expect just 1.40 years.[99] To appreciate the full social impact, we must next integrate this sort of disparity a second time over all defendants subject to racial bias, out of an approximate annual total of 20.7 million state criminal cases[100] and 70 thousand federal criminal cases.[101] And as Robert Abelson has demonstrated, even small percentages of variance explained might amount to huge impacts.[102]

B. The Civil Path

Now, we switch from the criminal to the civil path, and focus on the trajectory of an individual[103] bringing suit in a federal employment discrimination case, and how implicit bias might affect this process. First, the plaintiff, who is a member of a protected class, believes that her employer has discriminated against her in some legally cognizable way.[104] Second, after exhausting necessary administrative remedies,[105] the plaintiff sues in federal court. Third, the defendant tries to terminate the case before trial, via a motion to dismiss for failure to state a claim under Federal Rule of

---

99.   The simulation is available at <URL>. If in the simulation the effect size of race discrimination at each step is increased from r=.1 to r=.2, which is less than the average effect size of race discrimination effects found in the 2009 meta-analysis, see *supra* note 97, the ratio of expected years of sentence would increase to 3.11 years (Black) to 1.01 years (White).

100.   *See* R. LaFountain, R. Schauffler, S. Strickland, S. Gibson, & A. Mason, Examining the Work of State Courts: An Analysis of 2009 State Court Caseloads 3 (National Center for State Courts 2011) *available at* http://www.courtstatistics.org/FlashMicrosites/CSP/images/CSP2009.pdf (20.7 million).

101.   *See* Rachlinski et al., *supra* note 81, at 1202.

102.   *See* Robert P. Abelson, *A Variance Explanation Paradox: When a Little Is a Lot*, 97 Psychol. Bull. 129, 132 (1985) (explaining that a batting average of a .320 hitter or a .220 hitter predicts only 1.4 percent of the variance explained for a single at-bat producing either a hit or no-hit). Some discussion of this appears in Kang & Lane, *supra* note 2, at 489.

103.   We acknowledge that *Wal-Mart Stores, Inc. v. Dukes* made it much more difficult to certify large classes in employment discrimination cases. *See* Wal-Mart Stores, Inc. v. Dukes, 131 S. Ct. 2541, 2553-54 (2011) (holding that statistical evidence of gender disparities combined with a sociologist's analysis that Wal-Mart's corporate culture made it vulnerable to gender bias was inadequate to show that members of the putative class have a "common claim" for purposes of class certification under Federal Rule of Civil Procedure (FRCP) 23(b)).

104.   For example, in a Title VII "disparate treatment" cause of action, the plaintiff must demonstrate an adverse employment action "because of" the plaintiff's "race, color, religion, sex, or national origin." 42 U.S.C. § 2000(e)-2(a)(1) (2006). By contrast, in a Title VII "disparate impact" cause of action, the plaintiff challenges facially neutral policies that produce a disparate impact on protected populations. *See* Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971). We recognize that employment discrimination law is far more complex with different elements in different state and federal causes of action.

105.   The U.S. Equal Employment Opportunity Commission (EEOC) process is critical in practical terms because the failure to file a claim with the EEOC within the quite short statute of limitations (either 180 or 300 days depending on whether the jurisdiction has a state or local fair employment agency) or to timely file a court suit after a resort to the EEOC results in an automatic dismissal of the claim. However, neither EEOC inaction nor an adverse determination preclude private suit. *See* Charles Sullivan & Lauren Walter, Employment Discrimination Law and Practice § 12.03[B], at 672.

Civil Procedure (FRCP) 12(b)(6). Fourth, should that fail, the defendant moves for summary judgment under FRCP 56. Finally, should that motion also fail, the jury renders a verdict after trial. Again, at each of these stages,[106] implicit biases could potentially influence the outcome. To maintain a manageable scope of analysis, we focus on employer discrimination, pretrial adjudication, and jury verdict.

1.   Employer Discrimination

For many, the most interesting question is whether implicit bias helped cause the employer to discriminate against the plaintiff. There are good reasons to think that some negative employment actions are indeed caused by implicit biases in what tort scholars call a "but-for" sense. This "but-for" causation may be legally sufficient since Title VII and most state antidiscrimination statutes require only a showing that the plaintiff was treated less favorably "because of" a protected characteristic, such as race or sex.[107] But our objective here is not to engage the doctrinal[108] and philosophical questions[109] of whether existing antidiscrimination laws do or should recognize implicit bias-actuated discrimination. We also do not address what sorts of evidence should be deemed admissible when plaintiffs attempt to make such a case at trial.[110] Although those questions are

---

106.   As explained when we introduced the Criminal Path, the number of stages identified is somewhat arbitrary. We could have listed more or fewer stages.

107.   Section 703(a) of Title VII of the 1964 Civil Rights Act states that "[i]t shall be an unlawful employment practice for an employer to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual . . . because of [an] individual's race, color, religion, sex, or national origin." 42 U.S.C. § 2000(e)-2(a)(1) (2006).

108.   For discussion of legal implications, see Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161 (1995); Krieger & Fiske, *supra* note 2; Faigman et al., *supra* note 17.

109.   For a philosophical analysis, see Patrick S. Shin, *Liability for Unconscious Discrimination? A Thought Experiment in the Theory of Employment Discrimination Law*, 62 HASTINGS L.J. 67 (2010).

110.   For example, there is considerable disagreement on whether an expert should be allowed to testify that a particular case is an instance of implicit bias. This issue is part of a much larger debate regarding scientists' ability to make reasonable inferences about an individual case from group data. John Monahan and Laurens Walker first pointed out that scientific evidence often comes to court at two different levels of generality, one general and one specific. *See* Laurens Walker & John Monahan, *Social Frameworks: A New Use of Social Science in Law*, 73 VA. L. REV. 559 (1987). For instance, in a case involving the accuracy of an eyewitness identification, the general question might concern whether eyewitness identifications that are cross-racial are less reliable than same race identifications; the specific question in the case would involve whether the cross-racial identification in this case was accurate. Interested in social science evidence, Monahan and Walker referred to this as "social framework" evidence, though their fundamental insight regarding "frameworks" applies to all scientific evidence. In the context of implicit biases, then, general research amply demonstrates the phenomenon in the population. However, in the courtroom, the issue typically concerns whether a particular decision or action was a product of implicit bias.

As a scientific matter, knowing that a phenomenon exists in a population does not necessarily mean that a scientist can reliably say that it was manifest in a particular case. This has led to a debate as to whether experts should be limited to testifying only to the general phenomenon, or should be allowed to opine on whether a particular case is an instance of the general phenomenon.

critically important, our task is more limited—to give an empirical account of how implicit bias may potentially influence a civil litigation trajectory.

Our belief that implicit bias causes some employment discrimination is based on the following evidence. First, "tester studies" out in the field—which involve sending identical applicants or applications who differ only in race or gender—have generally uncovered discrimination. According to a summary by Mark Bendick and Ana Nunes, there have been "several dozen testing studies" in the past two decades, in multiple countries, focusing on discrimination against multiple demographic groups (including women, the elderly, and racial minorities).[111] These studies consistently reveal typical "net rates of discrimination" that range from 20 percent to 40 percent. In other words, in 20 to 40 percent of cases, employers treat subordinated groups (for example, racial minorities) worse than privileged groups (for example, Whites) even though the testers were carefully controlled to be identically qualified.

Second, although tester studies don't distinguish between explicit versus implicit bias, various laboratory experiments have found implicit bias correlations with discriminatory evaluations.[112] For example, Laurie Rudman and Peter Glick demonstrated that in certain job conditions, participants treat an "agentic" (self-promoting and competent) woman worse than an equally "agentic" man.[113] When the job description explicitly required the employee to be cooperative and to work well with others, participants rated the agentic female less hireable than the equally agentic

---

This is a complicated issue and scholars have weighed in on both sides of it. Opposing the use of expert testimony that a specific case is an instance of implicit bias, see David L. Faigman, Nilanjana Dasgupta, & Cecilia L. Ridgeway, *A Matter of Fit: The Law of Discrimination and the Science of Implicit Bias*, 59 HASTINGS L. J. 1389, 1394 (2008) ("The research does not demonstrate that an expert can validly determine whether implicit bias caused a specific employment decision."); and John Monahan, Laurens Walker & Gregory Mitchell, *Contextual Evidence of Gender Discrimination: The Ascendance of "Social Frameworks,"* 94 VA. L. REV. 1715, 1719 (2008) ("[Testimony] in which the expert witness explicitly linked general research findings on gender discrimination to specific factual conclusions ... exceeded the limitation on expert testimony established by the Federal Rules of Evidence and by both the original and revised proposal of what constitutes 'social framework' evidence."). Advocating allowing expert testimony that a particular case is an instance of some general phenomenon, see Susan T. Fiske & Eugene Borgida, *Standards for Using Social Psychological Evidence in Legal Proceedings*, 83 TEMPLE L. REV. 867, 876 (2011) ("Qualified social scientists who provide general, relevant knowledge and apply ordinary scientific reasoning may offer informal opinion about the individual case, but probabilistically.").

In the end, lawyers may be able to work around this dispute by using an expert to provide social framework evidence that identifies particular attributes that exacerbate biased decisionmaking, then immediately calling up another witness who is personally familiar with the defendant's work environment and asking that witness whether each of those particular attributes exists.

111.   *See* Marc Bendick Jr. & Ana P. Nunes, *Developing the Research Basis for Controlling Bias in Hiring* (forthcoming J. Social Issues 2012) (draft on file with authors).

112.   *See id.* at 11.

113.   Laurie A. Rudman & Peter Glick, *Prescriptive Gender Stereotypes and Backlash Toward Agentic Women*, 57 J. SOC. ISSUES 743, 757 (2001). "Agentic" was signaled by a life philosophy essay and canned answers to a videotaped interview that emphasized self-promotion and competence. *See id.* at 748. Agentic was contrasted against "androgynous," which added the characteristics of interdependence and cooperation. *Id.*

male.[114]   Probing deeper, the researchers identified that the participants penalized the female candidate for lack of social skills, not incompetence.[115]   Explicit bias measures did not correlate with the rankings; however, an implicit gender stereotype (associating women as more communal than agentic)[116] did correlate negatively with the social skills ratings.   In other words, the higher the implicit gender stereotype, the lower the social skills evaluation.[117]

   Third, field experiments have provided further confirmation under real-world conditions.   The studies by Marianne Bertrand and Sendhil Mullainathan demonstrating discrimination in callbacks on the basis of the names on comparable resumes have received substantial attention in the popular press as well as in law reviews.[118]   These studies found that for equally qualified—indeed, otherwise identical candidates—firms call back "Emily" more often than "Lakisha."   Less attention has been paid to Dan-Olof Rooth's extensions of this work, which found similar callback discrimination but also found correlations between implicit stereotypes and the discriminatory behavior.[119]   Rooth has found these correlations not only with implicit stereotypes about ethnic groups (Swedes versus Arab-Muslims) but also implicit stereotypes about the obese.[120]

   Because implicit bias in the *courtroom* is our focus, we will not attempt to offer a comprehensive summary of the scientific research as applied to the implicit bias in the *workplace*.[121]   We do, however, wish briefly to high-

---

114.   The difference was *M*=2.84 versus *M*=3.52 on a 5 point scale (p<0.05).   *See id.* at 753. No gender bias was shown when the job description was "masculine" and did not call for cooperative behavior.   Also, job candidates that were engineered to be androgynous—in other words, to show both agentic and cooperative traits—were treated the same regardless of gender. *See id.*

115.   *See id.* at 753–54.

116.   The "agentic" stereotype was captured by word stimuli such as "independent," "autonomous," and "competitive."   The "communal" stereotype was captured by words such as "communal," "cooperative," and "kinship."   *See id.* at 750.

117.   *See id.* at 756 (*r*=−0.49, *p*<0.001).   For further description of the study in the law reviews, see Kang, *supra* note 43, at 1517–18.

118.   Marianne Bertrand & Sendhil Mullainathan, *Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination*, 94 Am. Econ. Rev. 991 (2004).   A search of the TP-ALL database in Westlaw on December 10, 2011 revealed ninety-six hits.

119.   Dan-Olof Rooth, *Automatic Associations and Discrimination in Hiring: Real World Evidence*, 17 Labour Econ. 523 (2009) (finding that implicit stereotypes, as measured by the IAT, predicted differential callbacks of Swedish-named versus Arab/Muslim-named resumes).   A one standard deviation increase in implicit stereotype produced almost a 12 percent decrease in the probability that an Arab/Muslim candidate received an interview.

120.   Jens Agerström & Dan-Olof Rooth, *The Role of Automatic Obesity Stereotypes in Real Hiring Discrimination*, 96 J. Applied Psychol. 790 (2011) (finding that hiring managers (N=153) holding more negative IAT-measured automatic stereotypes about the obese were less likely to invite an obese applicant for an interview).

121.   Thankfully, many of these studies have already been imported in the legal literature. For review of the science, *see* Kang & Lane, *supra* note 2, at 484-85 (discussing evidence of racial bias in how actual managers sort resumes and of correlations between implicit biases, as measured by the IAT and differential callback rates).

light lines of research—variously called "constructed criteria," "shifting standards," or "casuistry"—that emphasize the *malleability of merit*. We focus on this work because it has received relatively little coverage in the legal literature and may help explain how complex decisionmaking with multiple motivations occurs in the real world.[122] Moreover, this phenomenon may influence not only the defendant (accused of discrimination) but also the jurors who are tasked to judge the merits of the plaintiff's case.

Broadly speaking, this research demonstrates that people frequently engage in motivated reasoning[123] in selection decisions that we justify by changing merit criteria on the fly, often without conscious awareness. In other words, as between two plausible candidates that have different strengths and weaknesses, we first choose the candidate we like—a decision that may well be influenced by implicit factors—and then justify that choice by molding our "merit" standards accordingly.

Let's make this point more concrete. In one experiment, Eric Luis Uhlmann and Geoffrey Cohen asked participants to evaluate two finalists for police chief—one male, the other female.[124] One candidate's profile signaled "book smart," the other's profile signaled "streetwise," and the experimental design varied which profile attached to the woman and which to the man. Regardless of which attributes the male candidate featured, participants favored the male candidate and articulated their hiring criteria accordingly. For example, education ("book smarts") was considered more important when the man had it.[125] Surprisingly, even the attribute of being "family oriented" and having children was deemed more important when the man had it.[126]

Michael Norton, Joseph Vandello, and John Darley have made similar findings, again in the domain of gender.[127] Participants were put in the role of manager of a construction company who had to hire a high-level employee. One candidate's profile signaled more education; the other's profile signaled more experience. Participants ranked these candidates (and three other filler candidates), and then explained their decisionmaking by writing down "what was most important in determining [their] decision."[128]

---

122.   One recent exception is Rich, *supra* note 22.

123.   For discussion of motivated reasoning in organization contexts, see Sung Hui Kim, *The Banality of Fraud: Re-Situating the Inside Counsel as Gatekeeper*, 74 FORDHAM L. REV. 983, 1029–34 (2005).

124.   *See* Eric Luis Uhlmann & Geoffrey L. Cohen, *Constructed Criteria: Redefining Merit to Justify Discrimination*, 16 PSYCHOL. SCI. 474, 475 (2005).

125.   *See id.* at 475 (M=8.27 with education versus M=7.07 without education, on a 11 point scale; p=0.006; d=1.02).

126.   *See id.* (M=6.21 with family traits versus 5.08 without family traits; p=0.05; d=0.86).

127.   Michael I. Norton, et al., *Casuistry and Social Category Bias*, 87 J. PERSONALITY & SOC. PSYCHOL. 817 (2004).

128.   *Id.* at 820.

In the control condition, the profiles were given with just initials (not full names) and thus the test subjects could not assess their gender. In this condition, participants preferred the higher educated candidate 76 percent of the time. In the two experimental conditions, the profiles were given names that signaled gender, with the man having higher education in one condition and the woman having higher education in the other. When the man had higher education, the participants preferred him 75 percent of the time. In sharp contrast, when the woman had higher education, only 43 percent of the participants preferred her.[129]

The fact of discrimination isn't as interesting as *how* the discrimination was justified. In the control condition and the man-has-more-education condition, the participants ranked education as more important than experience about half the time (48 percent and 50 percent).[130] By contrast, in the woman-has-more-education condition, only 22 percent ranked education as more important than experience.[131] In other words, what counted as merit was redefined, in real-time, to justify hiring the man.

Was this weighting done consciously, as part of a strategy to manipulate merit in order to provide a cover story for decisionmaking caused and motivated by explicit bias? Or was merit refactored in a more automatic, unconscious, dissonance-reducing rationalization, which would be more consistent with an implicit bias story? Norton and colleagues probed this causation question in another series of experiments, in the context of race and college admissions.[132] In a prior study, they had found that Princeton undergraduate students shifted merit criteria—the relative importance of GPA versus number of AP classes taken—to select the Black applicant over the White applicant who shared the same cumulative SAT score. To see whether this casuistry was explicit and strategic or implicit and automatic, they ran another experiment in which participants merely rated admissions criteria in the abstract without selecting a candidate for admission.

Participants were simply told that they were participating in a study examining the criteria most important to college admissions decisions. They were given two sample resumes to familiarize themselves with potential criteria. Both resumes had equivalent cumulative SAT scores, but differed on GPA (4.0 versus 3.6) versus number of AP classes taken (9 versus 6). Both resumes also disclosed the applicant's race. In one condition, the White candidate had the higher GPA (and fewer AP classes); in the other condition, the African American had the higher GPA (and fewer AP

---

129.   *Id.* at 821.
130.   *Id.*
131.   *Id.*
132.   Michael I. Norton et al., *Mixed Motives and Racial Bias: The Impact of Legitimate and Illegitimate Criteria on Decision Making*, 12 PSYCHOL. PUB. POL'Y & L. 36, 42 (2006).

classes).[133] After reviewing the samples, the participants had to rank order eight criteria in importance, including GPA, number of AP classes, SAT scores, athletic participation, etc.

In the condition with the Black candidate having the higher GPA, 77 percent of the participants ranked GPA higher in importance than number of AP classes taken. By contrast, when the White candidate had the higher GPA, only 63 percent of the participants ranked GPA higher than AP. This change in the weighting happened even though the participants did not expect that they were going to make an admissions choice or to justify that choice. Thus these differences could not be readily explained in purely strategic terms, as methods for justifying a subsequent decision. According to the authors, "[t]hese results suggest not only that it is possible for people to reweight criteria deliberately to justify choices but also that decisions made under such social constraints can impact information processing even prior to making a choice. This suggests that the bias we observed is not simply post hoc and strategic but occurs as an organic part of making decisions when social category information is present."[134]

The ways that human decisionmakers may adjust criteria in real-time, in subtle ways, to modify their judgments of merit, has significant salience for thinking about the ways that implicit bias may potentially influence employment decisions. In effect, bias can influence decisions in ways contrary to the standard, seemingly commonsensical, model. The conventional legal model describes behavior as a product of discrete and identifiable motives. This research suggests, however, that implicit motivations might influence behavior and that we then rationalize those decisions after-the-fact. Hence, some employment decisions might be motivated by implicit bias but rationalized post hoc on the basis of nonbiased criteria. This process of reasoning from behavior to motives, as opposed to the folk-psychology assumption that the arrow of direction is from motives to behavior, is, in fact, consistent with a large body of contemporary psychological research.[135]

2. Pretrial Adjudication: 12(b)(6)

As soon as the complaint is filed, the defendant will try to dismiss as many of the claims in the complaint as possible. Before recent changes in pleading, a motion to dismiss a complaint under FRCP 8 and 12(b)(6) was decided under the relatively lax standard of *Conley v. Gibson*. Under *Con-*

---

133.   *See id.*
134.   *Id.* at 46–47. This does not, however, fully establish that these differences were the result of implicit views rather than explicit ones. Even if test subjects did not expect to have to make admissions determinations, they might consciously select criteria that they believed favored one group over another.
135.   *See generally* TIMOTHY D. WILSON, STRANGERS TO OURSELVES: DISCOVERING THE ADAPTIVE UNCONSCIOUS (2002).

ley, all factual allegations made in the complaint were assumed to be true. As such, the court's task was simply to ask whether "it appears beyond doubt that the plaintiff can prove no set of facts in support of his claim."[136]

Starting in *Bell Atlantic Corp. v. Twombly*,[137] which addressed complex antitrust parallel conduct claims, and further developed in *Ashcroft v. Iqbal*,[138] which addressed civil rights actions based on racial and religious discrimination post-9/11, the U.S. Supreme Court abandoned the *Conley* standard. District courts must now, first, throw out factual allegations made in the complaint if they are merely conclusory.[139] Second, courts must make a "plausibility" decision based on the information before them. In the *Iqbal* case itself, the Supreme Court held that because of an "obvious alternative explanation"[140] of earnest national security responses, purposeful racial or religious "discrimination is not a plausible conclusion."[141]

How are courts supposed to decide what's "Twom-bal"[142] plausible when the motion to dismiss happens before discovery, especially in civil rights cases in which the defendant holds the key information? According to the Court, "[d]etermining whether a complaint states a plausible claim for relief will . . . be a context-specific task that requires the reviewing court to draw on its judicial experience and common sense."[143]

And when judges turn to their judicial experience and common sense, what will this store of knowledge tell them about whether some particular comment or act happened, and whether such behavior evidences legally cognizable discrimination? Decades of social psychological research demonstrate that our impressions are driven by the interplay between categorical (general to the category) and individuating (specific to the member of the category) information. For example, in order to come to an impression about a Latina plaintiff, we reconcile general schemas for Latina workers with individualized data about the specific plaintiff. When we lack sufficient individuating information—which is largely the state of affairs at a motion to dismiss stage—we have no choice but to rely more heavily on our schemas.[144]

---

136.   Conley v. Gibson, 355 U.S. 41, 45–46 (1957).

137.   550 U.S. 544 (2007).

138.   129 S.Ct. 1937 (2009).

139.   *See id.* at 1951.

140.   *Id.* (quoting *Bell Atlantic Corp. v. Twombly*, 550 U.S. 544).

141.   *Id.* at 1952.

142.   *See In re Iowa Ready-Mix Concrete Antitrust Litig.*, 2011 WL 5547159 (N.D. Iowa Nov. 9, 2011) (slip op.) (referring to *Twombly-Iqbal* motion as "Twom-bal").

143.   Ashcroft v. Iqbal, 129 S.Ct. at 1940.

144.   These schemas also reflect cultural cognitions. *See generally* Donald Braman, *Cultural Cognition and the Reasonable Person*, 14 Lewis & Clark L. Rev. 1455 (2010); Dan M. Kahan, David A. Hoffman & Donald Braman, *Whose Eyes Are You Going to Believe?* Scott v. Harris *and the Perils of Cognitive Illiberalism*, 122 Harv. L. Rev. 837 (2009).

Moreover, consider what the directive to rely on common sense means in light of "social judgeability theory."[145] According to this theory, there are social rules that tell us when it's appropriate to judge someone. For example, suppose your fourth grade child told you that a new kid, Hannah, has enrolled in school, and that she receives free lunches. Your child then asks you whether you think she's smart. You'll probably decline to answer since you do not feel entitled to make that judgment. Without more probative information, you feel that you would only be crudely stereotyping her abilities based on her socioeconomic status. But what if the next day you volunteered in the classroom and spent twelve minutes observing Hannah interacting with a teacher trying to solve problems? Would you then feel that you had enough individuating information to come to some judgment?

This is precisely what John Darley and Paget Gross tested in a seminal experiment in 1983.[146] When participants only received economic status information, they declined to evaluate Hannah's intelligence as a function of her economic class. However, when they saw a twelve-minute videotape of the child answering a battery of questions, participants felt credentialed to judge the girl, consistent with stereotypes. What they didn't realize was that the individuating information in the videotape was purposefully designed to be ambiguous. So participants who were told that Hannah was rich interpreted the video as confirmation that she was smart. By contrast, participants who were told that Hannah was poor interpreted the same video as confirmation that she wasn't so bright.[147]

Vincent Yzerbyt and colleagues, who call this phenomenon "social judgeability," have produced further evidence of this effect.[148] If researchers told you that a person is either an archivist or a comedian, then asked you twenty questions about this person regarding degree of extroversion with the options of "True," "False," or "I don't know," how might you answer? What if in addition they manufactured an illusion that you were given individuating information—information about the specific individual and not just the category he or she belongs to—even though you actually didn't receive any such information?[149] This is precisely what Yzerbyt and colleagues did in the lab.

---

145.   *See infra* note 148.

146.   *See* John M. Darley & Paget H. Gross, *A Hypothesis-Confirming Bias in Labeling Effects*, 44 J. PERSONALITY & SOC. PSYCHOL. 20, 22–23 (1983).

147.   *See id.* at 24–25, 27–29.

148.   *See* Vincent Y. Yzerbyt et al., *Social Judgeability: The Impact of Meta-Informational Cues on the Use of Stereotypes*, 66 J. PERSONALITY & SOC. PSYCHOL. 48 (1994).

149.   This illusion was created by having participants go through a listening exercise, in which they were told to focus only on one speaker (coming through one ear of a headset) and ignore the other (coming through the other). They were later told that the speaker that they were told to ignore had in fact provided relevant individuating information. The truth was, however, that no such information had been given. *See id.* at 50.

They found that those operating under the illusion of individuating information were more confident in their answers in that they marked fewer questions with "I don't know."[150] They also found that those operating under the illusion gave more stereotype-consistent answers.[151] In other words, the illusion of being informed made the target "judgeable." Because the participants in fact had received no such individuating information, they tended to judge the person in accordance with their schemas, about archivists and comedians. Interestingly, "in the debriefings, subjects reported that they did not judge the target on the basis of a stereotype; they were persuaded that they had described a real person qua person."[152] Again, it's possible that they were concealing their explicitly embraced bias about archivists and comedians from probing researchers, but we think that it is more probable that implicit bias explains these results.

Social judgeability theory connects back to *Iqbal* in that the Supreme Court has altered the rules structuring the judgeability of plaintiffs and their complaints. Under *Conley*, judges were told not to judge without the facts and thus were supposed to allow the lawsuit to get to discovery unless no set of facts could state a legal claim. By contrast, under *Iqbal*, judges have been explicitly green-lighted to judge the plausibility of the plaintiff's claim based only on the minimal facts that can be alleged before discovery—and this instruction came in the context of a racial discrimination case. In other words, our highest court has entitled district court judges to make this judgment based on a quantum of information that may provide enough facts to render the claim "socially judgeable" but not enough facts to ground that judgment in much more than the judge's schemas. Just as Yzerbyt's illusion of individuating information entitled participants to judge in the laboratory, the express command of the Supreme Court may entitle judges to judge in the courtroom when they lack any well-developed basis to do so.

There are no field studies to test whether biases, explicit or implicit, influence how actual judges decide motions to dismiss actual cases. It's not clear that researchers could ever collect such information. All that we have are some preliminary data about dismissal rates pre- and post-*Iqbal* that are consistent with our analysis. Again, since *Iqbal* made dismissals easier, we should see an increase in dismissal rates across the board.[153]

---

150.   *See id.* at 51 (*M*=5.07 versus 10.13; *p*<0.003).
151.   *See id.* (*M*=9.97 versus 6.30, out of 1 to 20 point range; *p*<0.006).
152.   *Id.*
153.   In the first empirical study of *Iqbal*, Hatamayr sampled 444 cases under *Conley* (from May 2005 to May 2007) and 173 cases under *Iqbal* (from May 2009 to August 2009). *See* Patricia W. Hatamayr, *The Tao of Pleading: Do* Twombly *and* Iqbal *Matter Empirically?*, 59 AM. U. L. REV. 553, 597 (2010). She found that the general rate of complaint dismissal rose from 46 percent to 56 percent. *See id.* at 602 tbl.2. However, this finding was not statistically significant under a Pearson chi-squared distribution test examining the different dismissal rates for *Conley*, *Twombly*, and *Iqbal* for three results: grant, mixed, and deny.

More relevant to our hypothesis is whether certain types of cases experienced differential changes in dismissal rates. For instance, we would expect *Iqbal* to generate greater increases in dismissal rates for race discrimination claims than, say, contract claims. (One reason is that judges are likely to have stronger biases that plaintiffs in the former type of case have less valid claims than those in the latter. Another reason is that we might expect some kinds of cases to raise more significant concerns about asymmetric information than do others. In contracts disputes, both parties may have good information about most of the relevant facts even prior to discovery. In employment discrimination cases, plaintiffs may have good hunches about how they've been discriminated against, but prior to discovery they may not have access to the broad array of information in the employer's possession that may be necessary to turn the 'hunch' into something a judge finds plausible. Moreover, these two reasons potentially interact: the more gap-filling and inference that a judge has to engage in, in the absence of a well-developed evidentiary record, the more room there may be for explicit and implicit biases to structure the judge's assessment.)

Notwithstanding the lack of field studies on these issues, there's some evidentiary support for these differential changes in dismissal rates. For example, Patricia Hatamayr sorted a sample of cases before and after *Iqbal* into six major categories: contracts, torts, civil rights, labor, intellectual property, and all other statutory cases.[154] She found that in contract cases, the rate of dismissal did not change much from *Conley* (32 percent) to *Iqbal* (32 percent).[155] By contrast, for Title VII cases, the rate of dismissal increased from 42 percent to 53 percent.[156] Victor Quintanilla has collected more granular data by counting not Title VII cases generally but federal employment discrimination cases filed specifically by Black plaintiffs both before and after *Iqbal*.[157] He found an even larger jump. Under the *Conley* regime, only 20.5 percent of the motions to dismiss such cases were granted. By contrast, under the *Iqbal* regime, 54.6 percent of them were granted.[158] These data lend themselves to multiple interpretations and suffer from various confounds. So at this point, we can make only modest claims. We merely suggest that the dismissal rate data are consistent with our hypothesis that *Iqbal*'s plausibility standard poses a risk of increasing the impact of implicit biases at the 12(b)(6) stage.

If, notwithstanding the plausibility-based pleading requirements, the case gets past the motion to dismiss, then discovery will take place, after which defendants will seek summary judgment under FRCP 56. On the

---

154.   *See id.* at 591-93.
155.   *See id.* at 630 tbl.D.
156.   *See id.*
157.   *See* Victor D. Quintanilla, *Beyond Common Sense: A Social Psychological Study of* Iqbal*'s Effect on Claims of Race Discrimination*, Mich. J. of Race & L. (forthcoming 2011). Quintanilla counted both Title VII and 42 U.S.C. § 1981 cases.
158.   *See id.* tbl.1 (*p*<0.000).

one hand, this procedural posture is less subject to implicit biases than the motion to dismiss because more individuating information will have surfaced through discovery. On the other hand, the judge still has to make a judgment call whether any "genuine dispute as to any material fact"[159] remains. Similar decisionmaking dynamics are likely to be in play as we saw in the pleading stage, for a significant quantum of discretion remains. Certainly the empirical evidence that demonstrates how poorly employment discrimination claims fare on summary judgment is not inconsistent with this view, though, to be sure, myriad other explanations of these differences are possible (including, for example, doctrinal obstacles to reaching a jury).[160]

3. Jury Verdict

If the case gets to trial, evidence will be introduced on the merits of the claim. Sometimes the evidence will be physical objects, such as documents, emails, photographs, voice recordings, evaluation forms, and the like. The rest of it will be witness or expert testimony, teased out and challenged by lawyers on both sides. Is there any reason to think that jurors might interpret the evidence in line with their biases? In the criminal trajectory, we already learned of juror bias via meta-analyses as well as correlations with implicit biases. Unfortunately, we lack comparable studies in the civil context. What we offer are two sets of related arguments and evidence that speak to the issue: motivation to shift standards and performer preference.

a. Motivation to Shift Standards

Above, we discussed the potential malleability of merit determinations when judgments permit discretion, and learned how employer defendants might shift standards and reweight criteria when evaluating applicants and employees. Here, we want to recognize that a parallel phenomenon may affect juror decisionmaking. Suppose that a particular juror is White and that he identifies strongly with his Whiteness. Suppose further that the defendant is White and is being sued by a racial minority. The accusation of illegal and immoral behavior threatens the status of the juror's (racial) ingroup. Anca Miron, Nyla Branscombe, and Monica Biernat have demonstrated that this threat to the ingroup can motivate people to shift

---

159.   FED R. CIV. P. 56(a).
160.   *See, e.g.,* Charlotte L. Lanvers, *Different Federal Court, Different Disposition: An Empirical Comparison of ADA, Title VI, Race and Sex, and ADEA Employment Dispositions in the Eastern District of Pennsylvania and the Northern District of Georgia*, 16 CORNELL J. L. & POL'Y 381, 395 (2007); Theodore Eisenberg & Charlotte Lanvers, *Summary Judgment Rates Over Time, Across Case Categories, and Across Districts: An Empirical Study of Three Large Federal Districts* (May 28, 2008), ssrn.com/abstract=1138373 (finding that civil rights cases, and particularly employment discrimination cases, have a consistently higher summary judgment rate than noncivil rights cases).

standards in a direction that shields the ingroup from ethical responsibility.[161]

Miron and colleagues asked White undergraduates at the University of Kansas to state how strongly they identified with America.[162] Then they were asked various questions about America's relationship to slavery and its aftermath. These questions clumped into three categories (or constructs): judgments of harm done to Blacks,[163] standards of injustice,[164] and collective guilt.[165] Having measured these various constructs, the researchers looked for relationships among them. Their hypothesis was that the greater the self-identification with America, the higher the standards would be before being willing to call America racist or otherwise morally blameworthy (that is, the participants would set higher "confirmatory standards"). They found that White students who strongly identified as American set higher standards for injustice (that is, they wanted more evidence before calling American unjust);[166] they thought less harm was done by slavery;[167] and as a result they felt less collective guilt compared to other White students who identified less with America.[168] In other words, their attitudes toward America were correlated with the quantum of evidence they required to reach a judgment that America had been unjust.

In a subsequent study, Miron et al. tried to find evidence of causation, not merely correlation. They did so by experimentally manipulating national identification by asking participants to recount situations in which they felt similar to other Americans (evoking greater identification with fellow Americans) or different from other Americans (evoking less identification with fellow Americans).[169] Those who were experimentally made to feel *less* identification with America subsequently reported very different standards of justice and collective guilt compared to others made to feel *more* identification with America. Specifically, participants in the low identification condition set lower standards for calling something unjust;

---

161. Anca M. Miron, Nyla R. Branscombe, & Monica Biernat, *Motivated Shifting of Justice Standards*, 36 PERSONALITY SOC. PSYCHOL. BULL. 768, 769 (2010).

162. The participants were all American citizens. The question asked was, "I feel strong ties with other Americans." *Id.* at 771.

163. A representative question was, "How much damage did Americans cause to Africans?" on a very little (1) to very much (7) Likert scale. *Id.* at 770.

164. "Please indicate what percentage of Americans would have had to be involved in causing harm to Africans for you to consider the past United States a racist nation" on a scale of 0–10 percent, 10–25 percent, up to 90–100 percent. *Id.* at 771.

165. "I feel guilty for my nation's harmful past actions toward African Americans" on a "strongly disagree" (1) to "strongly agree" (9) Likert scale. *Id.*

166. *See id.* at 772 tbl.I ($r=0.26$, $p<0.05$).

167. *See id.* at 772 tbl.I ($r=-0.23$, $p<0.05$).

168. *See id.* at 772 tbl.I ($r=-0.21$, $p<0.05$). Using structural equation modeling, the researchers found that "standards of injustice" fully mediated the relationship between group identification and judgments of harm; also "judgments of harm" fully mediated the effect of standards on collective guilt. *See id.* at 772–73.

169. The manipulation was successful. *See id.* at 773 ($p<0.05$, $d=0.54$.).

they evaluated slavery's harms as higher; and felt more collective guilt. By contrast, participants in the high identification condition set higher standards for calling something unjust (that is, they required more evidence); they evaluated slavery's harms as less severe; and they felt less guilt.[170] In other words by experimentally manipulating how much people identified with their ingroup (in this case "American"), researchers could shift the justice standard that participants deployed to judge their own ingroup for harming the outgroup.

Our evidentiary standards for jurors are specific but vague; in a civil case, a juror must find for the plaintiff by a preponderance of the evidence. The question is how a juror operationalizes that standard—just how much evidence does she require for believing that this standard has been met? These studies show how our assessments of evidence—of how much is enough—are themselves potentially malleable. One potential source of malleability is, according to this research, a desire (most likely implicit) to protect one's ingroup status. If a juror strongly identifies with the defendant employer as part of the same ingroup—racially or otherwise—the juror may shift standards of proof upwards in response to attack by an outgroup plaintiff. In other words, jurors who implicitly perceive an ingroup threat may require more evidence to be convinced of the defendant's harmful behavior than they would in an otherwise identical case that did not relate to their own ingroup. Ingroup threat is simply an example of this phenomenon; the point is that implicit biases may impact jurors by affecting how they implement ambiguous decision criteria regarding both the quantum of proof and how they make inferences from ambiguous pieces of information.

b.   Performer Preference

Jurors will often receive evidence and interpretive cues from "performers" at trial, by which we mean the cast of characters in the courtroom who jurors see, such as the judge, lawyers, parties, and witnesses. These various performers are playing roles of one sort or another. And, it turns out that people tend to have stereotypes about the ideal employee or worker that vary depending on the segment of the labor market. For example, in high-level professional jobs and leadership roles, the "ideal employee" is often a White man.[171] When the actual performer does not fit the ideal type, people may evaluate the performance more negatively.

---

170.   In standards for injustice, M=2.60 versus 3.39; on judgments of harm, M=5.82 versus 5.42; on collective guilt, M=6.33 versus 4.60. All differences were statistically significant at p=0.05 or less. *See id.* at 773.

171.   *See, e.g.*, Alice H. Eagly & Steven J. Karau, *Role Congruity Theory of Prejudice Toward Female Leaders*, 109 PSYCHOL. REV. 573 (2002); Alice H. Eagly, Steven J. Karau, & Mona G. Makhijani, *Gender and the Effectiveness of Leaders: A Meta-Analysis*, 117 PSYCHOL. BULL. 125 (1995). *See also* Shelley J. Correll et al., *Getting a Job: Is There a Motherhood Penalty?*, 112 AM. J. SOC. 1297

One study by Jerry Kang, Nilanjana Dasgupta, Kumar Yogeeswaran, and Gary Blasi found just such performer preference with respect to lawyers, as a function of race.[172] Kang and colleagues took jury-eligible participants from Los Angeles and measured both their explicit and implicit beliefs about the ideal lawyer. They were especially curious whether participants had implicit stereotypes linking the ideal litigator with particular racial groups (White versus Asian American, litigator versus scientist). In addition to measuring their biases, the researchers had participants evaluate two depositions, which they heard via headphones and also read on screen. At the beginning of each deposition, participants were shown for five seconds a picture of the litigator conducting the deposition on a computer screen accompanied by his name. The race of the litigator was varied by name and photograph. Also, the deposition transcript identified who was speaking, which meant that participants repeatedly saw the attorneys' last names.[173]

The study discovered the existence of a moderately strong implicit stereotype associating "litigator" with "White" (IAT D=0.45);[174] this stereotype correlated with more favorable evaluations of the White lawyer (ingroup favoritism) in terms of his competence (r=0.32, p<.01), likeability (r=0.31, p<.01), and hireability (r=0.26, p<.05). These results were confirmed through hierarchical regressions. To appreciate the magnitude of the effect sizes, imagine a juror who has no explicit stereotype but a large implicit stereotype (IAT D=1) that the ideal litigator is White. On a 7-point scale, this juror would favor a White lawyer over an identical Asian American lawyer 6.01 to 5.65 in terms of competence; 5.57 over 5.27 in terms of likability, and 5.65 over 4.92 in terms of hireability.[175]

This study provides some evidence of racial discrimination against attorneys in the judgment of their performance of a basic deposition by potential jurors, as a function of jurors' implicit stereotypes. What does this have to do with how juries might decide employment discrimination cases? Of course minority defendants don't necessarily hire minority attorneys. That said, it's possible that minorities do hire minority attorneys at somewhat higher rates than nonminorities. But even more important, we hypothesize that similar processes might take place with how jurors evaluate not only attorneys, but also both parties and witnesses, as they perform

---

(2007); Joan Williams, Unbending Gender; Why Family and Work Conflict and What to do About It 213–17 (2000) (discussing how what counts as meritorious is designed around masculine norms).

    172.   *See* Jerry Kang et al., *Are Ideal Litigators White? Measuring the Myth of Colorblindness*, 7 J. Empirical Leg. Studies 886 (2010).

    173.   *See id.* at 892-99 (describing method and procedure, and identifying attorney names as "William Cole" or "Sung Chang").

    174.   *See id.* at 900. They also found strong negative implicit attitudes against Asian Americans (IAT D=0.62). *See id.*

    175.   These figures were calculated using the regression equations in *id.* at 902-03 nn. 25, 27.

their various roles at trial. To be sure, this study does not speak directly to credibility assessments, likely to be of special import at trial, but it does at least suggest that implicit stereotypes may affect judgment of "performances" in the courtroom.

We concede that our claims about implicit bias influencing jury decisionmaking in civil cases are somewhat speculative and not well quantified. Moreover, in the real world, certain institutional processes may make both explicit and implicit biases less likely to translate into behavior. For example, jurors must deliberate with other jurors, and sometimes the jury features significant demographic diversity, which seems to deepen certain types of deliberation.[176] Jurors also feel accountable[177] to the judge, who reminds them to adhere to the law and the merits. That said, for reasons already discussed, it seems implausible to think that current practices within the courtroom somehow magically burn away all jury biases, especially implicit biases, that jurors and judges are unaware of. That's why we seek improvements based on the best understanding of how people actually behave.

Thus far, we have canvassed much of the available evidence describing how implicit biases may influence decisionmaking processes in both criminal and civil cases. On the one hand, the research findings are substantial and robust. On the other hand, they provide only imperfect knowledge, especially about what is actually happening in the real world. Notwithstanding this provisional and limited knowledge, we strongly believe that these studies, in aggregate, suggest that implicit bias in the trial process is a problem worth worrying about. What, then, can be done? Based on what we know, how might we intervene to improve the trial process and potentially vaccinate decisionmakers against, or at least reduce, the influence of implicit bias?

III.   Interventions

Before we turn explicitly to interventions, we reiterate that there are many causes of unfairness in the courtroom, and our focus on implicit bias is not meant to deny other causes. That said, partly because of their invisibility, implicit biases are likely to be ignored even by the well intentioned and hence, good faith and awareness are not, by themselves, likely to prevent their operation. In Part II, we laid out the empirical case why we believe that implicit biases influence both criminal and civil case trajectories. We now identify interventions that build on an overlapping scientific and political consensus. If there are cost-effective interventions that are likely

---

176.   *See infra* text accompanying notes 224–228.

177.   *See, e.g.*, Jennifer S. Lerner & Philip E. Tetlock, *Accounting for the Effects of Accountability*, 125 Psychol. Bull. 255, 267–70 (1999).

to decrease the impact of implicit bias in the courtroom, we believe they should be adopted at least as forms of experimentation.

We are mindful of potential costs, including implementation and even overcorrection costs. But we are hopeful that these costs can be safely minimized. Moreover, the potential benefits to these improvements are both substantive and expressive. Substantively, they may increase actual fairness (by decreasing the impact of implicit biases); expressively, they may increase the appearance of fairness, by signaling the judiciary's thoughtful attempts to go beyond "cosmetic compliance."[178] Effort isn't always sufficient, but it ought to count for something.

A. Decrease the Implicit Bias

If implicit bias causes unfairness, one intervention strategy is to decrease the implicit bias itself. It would be delightful if explicit refutation would suffice. But abstract global self-commands to "Be fair!" don't much change implicit social cognitions. How then might we alter implicit associations between certain social groups on the one hand and negative attitudes or stereotypes on the other?[179] One potentially effective strategy is to expose ourselves to countertypical associations. In rough terms, if we have a negative attitude toward some group, we need exposure to members of that group to whom we would have a positive attitude. If we have a particular stereotype about some group, we need exposure to members of that group that don't feature those particular attributes.

These exposures can come through direct contact with countertypical human beings. For example, Nilanjana Dasgupta and Shaki Asgari tracked the implicit gender stereotypes held by women subjects both before and after a year of attending college.[180] One group of women attended a year of coed college; the other group attended a single-sex college. At the start of their college careers, the two groups had comparable amounts of implicit stereotypes against women. However, one year later, those who attended the women's college on average expressed no gender bias whereas the av-

---

178. In a 1999 survey by the National Center for State Courts, 47 percent of the American people doubted that African Americans and Latinos receive equal treatment in state courts; 55 percent doubted that non-English speaking people receive equal treatment. The appearance of fairness is a serious problem. *See* NATIONAL CENTER FOR STATE COURTS, HOW THE PUBLIC VIEWS THE STATE COURTS: A 1999 NATIONAL SURVEY 37 (1999), *available at* http://www.ncsconline.org/WC/Publications/Res_AmtPTC_PublicView CrtsPub.pdf. The term "cosmetic compliance" comes from Kimberly D. Krawiec, *Cosmetic Compliance and the Failure of Negotiated Governance*, 81 WASH. U. L.Q. 487 (2003).

179. For analysis of the nature versus nurture debate regarding implicit biases, see JERRY KANG, *Bits of Bias*, *in* IMPLICIT BIAS ACROSS THE LAW (Justin Levinson & Robert Smith, eds., forthcoming 2012).

180. *See* Nilanjana Dasgupta & Shaki Asgari, *Seeing Is Believing: Exposure to Counterstereotypic Women Leaders and Its Effect on the Malleability of Automatic Gender Stereotyping*, 40 J. EXPERIMANTAL SOC. PSYCHOL. 642, 649–54 (2004).

erage bias of those who attended the coed school increased.[181] By carefully examining differences in the two universities' environments, the researchers learned that it was exposure to countertypical women in the role of professors and university administrators that altered the implicit gender stereotypes of female college students.[182]

Nilanjana Dasgupta and Luis Rivera also found correlations between participants' self-reported numbers of gay friends and their negative implicit attitudes toward gays.[183] Such evidence gives further reason to encourage intergroup social contact by diversifying the bench, the courtroom (staff and law clerks), our residential neighborhoods, and friendship circles. That said, any serious diversification of the bench, the bar, and staff would take enormous resources, both economic and political. Moreover, these interventions might produce only modest results. For instance, Rachlinski et al. found that judges from an Eastern district that featured approximately half White judges and half Black judges had "only slightly smaller" implicit biases than the judges of a Western jurisdiction, which contained only two Black judges (out of forty-five total district court judges, thirty-six of them being White).[184] In addition, "debiasing" exposures would have to compete against the other daily real life exposures in the courtroom that "re-bias." For instance, Joshua Correll found that police officers who worked in areas with high minority demographics and violent crime showed more shooter bias.[185]

If increasing direct contact with a diverse but countertypical population isn't readily feasible, what about "vicarious" contact, which is mediated by images, videos, simulations, or even imagination, and does not require direct face-to-face contact?[186] Actually, the earliest studies on the malleability of implicit bias pursued just these strategies. For instance, Nilanjana Dasgupta and Anthony Greenwald showed that participants exposed vicariously, in a history questionnaire, to countertypical exemplars (for example, Black figures to whom we tend to have positive attitudes, such as Martin Luther King Jr. and White figures to whom we tend to have negative attitudes, such as Charles Manson) showed a substantial decrease in negative implicit attitudes toward African Americans.[187] These findings

---

181.   *See id.* at 651.

182.   *See id.* at 651-53.

183.   *See* Nilanjana Dasgupta & Luis M. Rivera, *From Automatic Antigay Prejudice to Behavior: The Moderating Role of Conscious Beliefs About Gender and Behavioral Control*, 91 J. PERSONALITY & SOC. PSYCHOL. 268, 270 (2006).

184.   *See* Rachlinski et al., *supra* note 81, at 1227.

185.   *See* Correll et al., *supra* note 48, at 1014 ("We tentatively suggest that these environments may reinforce cultural stereotypes, linking Black people to the concept of violence.").

186.   *See* Jerry Kang, *Cyber-Race*, 113 HARV. L. REV. 1130, 1166–67 (comparing vicarious versus direct experiences).

187.   Nilanjana Dasgupta & Anthony G. Greenwald, *On the Malleability of Automatic Attitudes: Combating Automatic Prejudice With Images of Admired and Disliked Individuals*, 81 J. PERSONALITY & SOC. PSYCH. 800, 807 (2001). The IAT effect changed nearly 50 percent as

are consistent with work done by Irene Blair, who has demonstrated that brief mental visualization exercises can also change scores on the IAT.[188]

In addition to exposing people to famous countertypical exemplars, implicit biases may be decreased by juxtaposing ordinary people with countertypical settings. For instance, Bernard Wittenbrink, Charles Judd, and Bernadette Park examined the effects of watching videos of African Americans situated either at a convivial outdoor barbecue or at a gang-related incident.[189] Situating African Americans in a positive setting produced lower implicit bias scores.

There are, to be sure, questions about whether this evidence directly translates into possible improvements for the courtroom.[190] But even granting numerous caveats, might it not be valuable to engage in some experimentation? In chambers and the courtroom buildings, photographs, posters, screen savers, pamphlets, and decorations ought to be used that bring to mind countertypical exemplars or associations for participants in the trial process. Since judges and jurors are differently situated, we can expect both different effects and implementation strategies. For example, judges would be exposed to such vicarious displays regularly as a feature of their workplace environment. By contrast, jurors would be exposed only during their typically brief visit to the court.[191] Especially for jurors, then, the goal is not anything as grand as fundamentally changing the underlying structure of their mental associations. Instead, the hope would be that by reminding them of countertypical associations, we might momentarily activate different mental patterns while in the courthouse and reduce the impact of implicit biases on their decisionmaking.[192]

To repeat, we recognize the limitations of our recommendation. Recent research has found much smaller debiasing effects from vicarious exposure than originally estimated.[193] Moreover, such exposures must compete against the flood of typical, schema-consistent exposures we are bom-

---

compared to the control (IAT effect *M*=78ms versus 174ms, *p*=0.01), and remained for over twenty-four hours.

188. Irene V. Blair, Jennifer E. Ma, & Alison P. Lenton, *Imagining Stereotypes Away: The Moderation of Implicit Stereotypes Through Mental Imagery*, 81 J. PERSONALITY & SOC. PSYCHOL. 828 (2001). *See generally* Irene V. Blair, *The Malleability of Automatic Stereotypes and Prejudice*, 6 PERSONALITY & SOC. PSYCHOL. REV. 242 (2002) (literature review).

189. *See* Bernd Wittenbrink et al., *Spontaneous Prejudice in Context: Variability in Automatically Activated attitudes*, 81 J. PERSONALITY & SOC. PSYCHOL. 815, 818–19 (2001).

190. How long does the intervention last? How immediate does it have to be? How much were the studies able to insure focus on the positive countertypical stimulus as opposed to in a courtroom where these "positives" would be amidst the myriad distractions of the trial?

191. *See* Kang, *supra* note 43, at 1537 (raising possibility of "debiasing booths" in lobbies for waiting jurors).

192. Rajees Sritharan & Bertram Gawronski, *Changing Implicit and Explicit Prejudice: Insights from the Associative-Propositional Evaluation Model*, 41 SOC. PSYCHOL. 113, 118 (2010).

193. *See* Jennifer A. Joy-Gaba & Brian A. Nosek, *The Surprisingly Limited Malleability of Implicit Racial Evaluations*, 41 SOC. PSYCHOL. 137, 141 (2010) (finding an effect size as compared to the original Dasgupta and Greenwald findings, *supra* note 187, that were approximately 70 percent smaller).

barded with from mass media. That said, we see little costs to these strategies even if they appear "cosmetic." There is no evidence, for example, that these exposures will be so powerful so as to overcorrect and produce net bias against Whites.

B.   Break the Link Between Bias and Behavior

Even if we can't remove the bias, perhaps we can alter decisionmaking processes so that these biases are less likely to translate into behavior. In order to keep this Article's scope manageable, we focus on the two key players in the courtroom: judges and jurors.[194]

1.   Judges

a.   Doubt One's Objectivity

Most judges view themselves as objective and especially talented at fair decisionmaking. For instance, Rachlinski et al. found in one survey that 97 percent of judges (thirty-five out of thirty-six) believed that they were in the top quartile on "avoid[ing] racial prejudice in decisionmaking"[195] relative to other judges attending the same conference. That is, obviously, mathematically impossible. (One is reminded of Lake Wobegon, where all of the children are above average.) In another survey, 97.2 percent of those administrative agency judges surveyed put themselves in the top half in terms of avoiding bias, again impossible.[196] Unfortunately, there's evidence that believing oneself to be objective puts us at particular risk for behaving in ways that belie our self-conception.

Eric Uhlmann and Geoffrey Cohen have demonstrated that when a person believes himself to be objective, such belief licenses him to act on his biases. In one study, they had participants choose either a candidate "Gary" or "Lisa" for the job of factory manager. Both candidate profiles, comparable on all traits, unambiguously showed strong organization skills but weak interpersonal skills.[197] Half the participants were primed to view themselves as objective.[198] The other half were left alone as control.

---

194.   Other important players obviously include staff, lawyers, and police. For a discussion of the training literature on the police and shooter bias, see Adam Benforado, *Quick on the Draw: Implicit Bias and the Second Amendment*, 89 OR. L. REV. 1, 46–48 (2010).

195.   *See* Rachlinski et al., *supra* note 81, at 1225.

196.   *See* Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, *The "Hidden Judiciary": An Empirical Examination of Executive Branch Justice*, 58 DUKE L.J. 1477, 1519 (2009).

197.   *See* Eric Luis Uhlmann & Geoffrey L. Cohen, *"I think it, therefore it's true": Effects of Self-Perceived Objectivity on Hiring Discrimination*, 104 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 207, 210-11 (2007).

198.   This was done simply by asking participants to rate their own objectivity. Over 88 percent of the participants rated themselves as above average on objectivity. *See id.* at 209. The participants were drawn from a lay sample (not just college students).

Those in the control condition gave the male and female candidates statistically indistinguishable hiring evaluations.[199] But those who were manipulated to think of themselves as objective evaluated the male candidate higher (M=5.06 versus 3.75, p=0.039, d=0.76).[200] Interestingly, this wasn't due to a malleability of merit effect, in which the participants reweighted the importance of either organizational skills or interpersonal skills in order to favor the man. Instead, the discrimination was caused by straight-out disparate evaluation, in which "Gary" was rated as more interpersonally skilled than "Lisa" by those primed to think themselves objective (M=3.12 versus 1.94, p=0.023, d=0.86).[201] In short, thinking oneself to be objective seems to lead one ironically to be less objective and more susceptible to biases. Judges should therefore remind themselves that notwithstanding their status, education, and the robe, they are human and fallible.

But is such a suggestion based on wishful thinking? Is there any evidence that education and reminders can actually help? There's some suggestive evidence from Emily Pronin, who has carefully studied the bias blindspot—the belief that others are biased but we ourselves are not.[202] In one study, Emily Pronin and Matthew Kugler had a control group of Princeton students read an article from *Nature* about environmental pollution. By contrast, the treatment group read an article allegedly published in *Science* that described various nonconscious influences on attitudes and behaviors.[203] After reading an article, the participants were queried about their own objectivity as compared to their university peers. Those in the control group revealed the predictable bias blindspot and thought that they suffered from less bias than their peers.[204] By contrast, those in the treatment group did not believe that they were more objective than their peers; moreover, their more modest self-assessments differed from those of the more confident control group.[205] These results suggest that learning about nonconscious thought processes can lead people to be more skeptical about their own objectivity.

---

199. *See id.* at 210-11 (*M*=3.24 for male candidate versus 4.05 for female candidate, *p*=0.21).

200. *See id.* at 211.

201. *See id.* Interestingly, the gender of the participants mattered. Female participants did not show the objectivity priming effect. *See id.*

202. *See generally* Emily Pronin, *Perception and Misperception of Bias in Human Judgment*, 11 Trends in Cognitive Sciences 37 (2007).

203. *See* Emily Pronin & Matthew B. Kugler, *Valuing Thoughts, Ignoring Behavior: The Introspective Illusion as a Source of the Bias Blind Spot*, 43 J. Experimental Soc. Psychol. 565, 574 (2007). The intervention article was 1643 words long, excluding references. *See id.* at 575.

204. *See id.* at 575 (M=5.29 where 6= "same" amount of bias as peers).

205. *See id.* For the treatment group, their self-evaluation of objectivity was M=5.88, not statistically significantly different from the score of 6, which meant having the "same" amount of bias as peers. Also, the self-reported objectivity of the treatment group (M=5.88) differed from the control group (M=5.29) in a statistically significant way, *p*=.01. *See id.*

b. Increase Motivation

Tightly connected to doubting one's objectivity is the strategy of increasing one's motivation to be fair.[206] Social psychologists generally agree that motivation is an important determinant of checking biased behavior.[207] Specific to implicit bias, Nilanjana Dasgupta and Luis Rivera found that participants who were consciously motivated to be egalitarian did not have their antigay implicit attitudes translate into biased behavior toward a gay person. By contrast, for those lacking such motivations, strong antigay implicit attitudes predicted more biased behavior.[208]

A powerful way to increase judicial motivation is to gain actual scientific knowledge about implicit social cognitions. In other words, judges should be internally persuaded that a genuine problem exists. This education and awareness can be done through self-study as well as more official judicial education. Such education is already taking place, although mostly in an ad hoc fashion.[209] The most organized intervention has come through the National Center for State Courts (NCSC). The NCSC organized a three-state pilot project in California, Minnesota, and North Dakota to teach judges and court staff about implicit bias.[210] It used a combination of written materials, videos, resource websites, Implicit Association Tests, and online lectures from subject matter experts to provide the knowledge. Questionnaires before and after each educational intervention provided an indication of program effectiveness.

Although increased knowledge of the underlying science is a basic objective of an implicit bias program, the goal is not to send judges back to college for a crash course in Implicit Psychology 101. Rather, it is to persuade judges, on the merits, to recognize implicit bias as a potential problem, which in turn should increase motivation to adopt sensible countermeasures. Did the NCSC projects increase recognition of the problem and encourage the right sorts of behavioral changes? The only evidence we have is limited, via voluntary self-reports subject to obvious selection biases.

---

206. For a review, see Margo J. Monteith et al., *Schooling the Cognitive Monster: The Role of Motivation in the Regulation and Control of Prejudice*, 3 SOC. & PERSON. PSYCHOL. COMPASS 211 (2009).

207. *See* Russell H. Fazio & Tamara Towles-Schwen, *The MODE Model of Attitude-Behavior Processes*, *in* DUAL PROCESS THEORIES IN SOCIAL PSYCHOLOGY 97 (Shelly Chaiken & Yaacov Trope eds., 1999).

208. *See* Dasgupta & Rivera, *supra* note 183, at 275.

209. Several of the authors of this Article have spoken to judges on the topic of implicit bias.

210. *See* NCSC, Helping Courts Address Implicit Bias: Resources for Education (forthcoming 2012).

For example, in California, training emphasized a documentary on the neuroscience of bias.[211] Before and after watching the documentary, participants were asked to what extent they thought "a judge's decisions and court staff's interaction with the public can be unwittingly influenced by unconscious bias toward racial/ethnic groups?" Before viewing the documentary, approximately 16 percent chose "rarely-never," 55 percent "occasionally," and 30 percent "most-all." After viewing the documentary, 1 percent chose "rarely-never," 20 percent "occasionally," and 79 percent "most-all."

Relatedly, participants were asked whether they thought implicit bias could have an impact on behavior even if a person lacked explicit bias. Before viewing the documentary, approximately 9 percent chose "rarely-never," 45 percent "occasionally," and 45 percent "most-all;" after viewing the documentary, 1 percent chose "rarely-never," 14 percent "occasionally," and 84 percent "most-all." These statistics provide some evidence that the California documentary increased awareness of the problem of implicit bias. The qualitative data, in the form of write-in comments[212] support this interpretation.

What about adoption of behavioral countermeasures? Because there wasn't a specific set of recommendations at the time of training, there was no attempt to measure behavioral changes. All that we have are self-reports that speak to the issue. For instance, participants were asked to agree or disagree with the statement, "I will apply the course content to my work." In California, 90 percent (N=60) reported that they agreed or strongly agreed.[213] In North Dakota (N=32), 97 percent reported that they agreed or strongly agreed.[214] Three months later, there was a follow-up survey given to the North Dakota participants, with only fourteen persons replying. In that survey, 77 percent of those who responded stated that they had made efforts to reduce the potential impact of implicit bias.[215] In

---

211. *See The Neuroscience and Psychology of Decisionmaking*, ADMIN. OFF. COURTS, EDUC. DIV. (Mar. 29, 2011), http://www2.courtinfo.ca.gov/cjer/aoctv/dialogue/neuro/index.htm. This program was broadcast on the Judicial Branch's cable TV station and made available streaming on the Internet.

212. Comments included: "raising my awareness of prevalence of implicit bias," "enlightened me on the penetration of implicit bias in everyday life, even though I consciously strive to be unbiased and assume most people try to do the same," and "greater awareness—I really appreciated the impressive panel of participants; I really learned a lot, am very interested." *See* NCSC, *supra* note 210, at ___.

213. *See id.* at 10.

214. *See id.* at 18. Minnesota answered a slightly different question: 81 percent gave the program's applicability a medium high to high rating.

215. *See id.* at 20. The strategies that were identified included: "concerted effort to be aware of bias," "I more carefully review my reasons for decisions, likes, dislikes, and ask myself if there may be bias underlying my determination," "Simply trying to think things through more thoroughly," "Reading and learning more about other cultures," and "I have made mental notes to myself on the bench to be more aware of the implicit bias and I've re-examined my feelings to see if it is because of the party and his/her actions vs. any implicit bias on my part."

sum, the findings across all three pilot programs suggest that education programs can increase motivation and encourage judges to engage in some behavioral modifications. Given the limitations of the data (for example, pilot projects with small numbers of participants, self-reports, self-selection, and limited follow-up results), additional research is needed to confirm these promising but preliminary results.

From our collective experience, we also recommend the following tactics. First, training should commence early, starting with new judge orientation when individuals are likely to be most receptive. Second, training should not immediately put judges on the defensive, for instance, by accusing them of concealing explicit bias. Instead, one can start the conversation with other types of decisionmaking errors and cognitive biases, such as anchoring, or less-threatening biases, such as the widespread preference for the youth over the elderly as revealed by IATs. Third, judges should be encouraged to take the IAT or other measures of implicit bias. Numerous personal accounts have reported how the discomfiting act of taking the IAT in and of itself motivates action. And researchers are currently studying the specific behavioral and social cognitive changes that take place through such self-discovery. That said, we do not recommend that such tests be strictly mandatory because the feeling of resentment and coercion is likely to counter the benefits of increased self-knowledge. Moreover, judges should never be expected to disclose their personal result.

c. Improve Conditions of Decisionmaking

Implicit biases function automatically. One way to counter them is to engage in effortful, deliberative processing.[216] But when decisionmakers are short on time or under cognitive load, they lack the resources necessary to engage in such deliberation. Accordingly, we encourage judges to take special care when they must respond quickly and to try to avoid making snap-judgments whenever possible. We recognize that judges are under enormous pressures to clear ever-growing dockets. That said, it is precisely under such work conditions that judges need to be especially on guard against their biases.

There is also evidence that certain elevated emotional states, either positive or negative, can prompt more biased decisionmaking. For exam-

---

216. There are also ways to deploy more automatic countermeasures. In other words, one can teach one's mind to respond not reflectively but reflexively, by automatically triggering goal-directed behavior through internalization of certain "if-then" responses. These countermeasures function implicitly and even under conditions of cognitive load. *See generally* Saaid A. Mendoza et al., *Reducing the Expression of Implicit Stereotypes: Reflexive Control Through Implementation Intentions*, 36 Personality & Soc. Psychol. Bull. 512, 514–15, 520 (2010); Monteith et al., *supra* note 206, at 218–21 (discussing "bottom-up" correction versus "top-down").

ple, a state of happiness seems to increase stereotypic thinking,[217] which can be countered when individuals are held accountable for their judgments. Of greater concern might be feelings of anger, disgust, or resentment toward certain social categories. If the emotion is consistent with the stereotypes or anticipated threats associated with that social category, then those negative emotions are likely to exacerbate implicit biases.[218]

In sum, judges should try to achieve the conditions of decisionmaking that allow them to be mindful and deliberative, without huge emotional swings. All the while, judges should never assume that they have ever reached anything like objectivity.

d. Count

Finally, we encourage judges and judicial institutions to count. One frequently offered mechanism for reducing bias has been to attempt to increase accountability, which has been shown to decrease the influence of bias. But, how can the behavior of trial court judges be held accountable if biased decisionmaking is itself difficult to detect? If judges do not seek out the information that could help them see their own potential biases, they become more difficult to correct as well. Just as trying to lose or gain weight without a scale is challenging, judges should engage in more quantified self-analysis and seek out and assess patterns of behavior that cannot be recognized in single decisions. Judges need to count.

The comparison we want to draw is with professional umpires and referees. Statistical analyses by behavioral economists have discovered various biases, including ingroup racial biases, in the decisionmaking of professional sports judges. Joseph Price and Justin Wolfers found racial ingroup biases in NBA Referees' calling fouls;[219] Christopher Parsons and colleagues found ingroup racial bias in MLB umpires' calling strikes.[220]

---

217.   *See* Galen V. Bodenhausen et al., *Happiness and Stereotypic Thinking in Social Judgment*, 66 J. PERSONALITY & SOC. PSYCHOL. 621 (1994).

218.   *See* Nilanjana Dasgupta et al., *Fanning the Flames of Prejudice: The Influence of Specific Incidental Emotions on Implicit Prejudice*, 9 EMOTION 585 (2009). The researchers found that implicit bias against gays and lesbians could be increased more by making participants feel "disgust" as compared to "anger." *See id.* at 588. Conversely, they found that implicit bias against Arabs could be increased more by making participants feel "angry" rather than "disgusted." *See id.* at 589. *See also* David DeSteno. et al., *Prejudice From Thin Air: The Effect of Emotion on Automatic Intergroup Attitudes*, 15 PSYCHOL. SCIENCE 319 (2004).

219.   Joseph Price & Justin Wolfers, *Racial Discrimination Among NBA Referees*, 125 Q. J. ECON. 1859, 1885 (2010) ("We find that players have up to 4% fewer fouls called against them and score up to 2½% more points on nights in which their race matches that of the refereeing crew. Player statistics that one might think are unaffected by referee behavior [for example, free throw shooting] are uncorrelated with referee race. The bias in foul-calling is large enough so that the probability of a team winning is noticeably affected by the racial composition of the refereeing crew assigned to the game.").

220.   Christopher A. Parsons et al., *Strike Three: Discrimination, Incentives, and Evaluation*, 101 AM. ECON. REV. 1410, 1433 (2011) ("Pitches are slightly more likely to be called strikes when the umpire shares the race/ethnicity of the starting pitcher, an effect that is observable only when

These discoveries were only possible because professional sports leagues count performance, including referee performance, in a remarkably granular and comprehensive manner.

Although NBA referees and MLB umpires make more instantaneous calls than judges, judges do regularly make quick judgments on motions, objections, and the like. In these contexts, often, judges can't slow down. So, it makes sense to count their performance in domains such as bail, probable cause, and preliminary hearings.

We recognize that such counting may be difficult for individual judges who lack both the quantitative training and the resources to track their own performance statistics. That said, even amateur, basic counting, with data collection methods never intended to make it into a peer-reviewed journal, might reveal surprising outcomes. Of course, the most useful information will require an institutional commitment to counting across multiple judges and will make use of appropriately sophisticated methodologies. The basic objective is to create a negative feedback loop in which individual judges and the judiciary writ large are given the corrective information necessary to know how they are doing and to be motivated to make changes if they find evidence of biased performances. It may be difficult to correct biases even when we do know about them, but it's virtually impossible to correct them if they remain invisible.

2. Jurors

a. Jury Selection and Composition

Individual screen. As for jurors, one obvious way to break the link between bias and unfair decisions is to keep biased persons off the jury. Since everyone has implicit biases of one sort or another, the more precise goal would be to screen out those with excessively high biases that are relevant to the case at hand. This is, of course, precisely one of the purposes of voir dire although the interrogation process was designed to ferret out concealed explicit bias, not implicit bias.

One might reasonably ask whether potential jurors should be individually screened for implicit bias via some instrument such as the IAT. But the leading scientists in implicit social cognition recommend against using the test as an individually diagnostic measure. One reason is that although the IAT has enough test-retest reliability to provide useful research information about human beings generally, its reliability is sometimes below

---

umpires' behavior is not well monitored. The evidence also suggests that this bias has substantial effects on pitchers' measured performance and games' outcomes. The link between the small and large effects arises, at least in part, because pitchers alter their behavior in potentially discriminatory situations in ways that ordinarily would disadvantage themselves (such as throwing pitches directly over the plate.").

what we would like for individual assessments.[221]   (That said, readers should understand that "the IAT's properties approximately resemble those of sphygmomanometer blood pressure (BP) measures that are used to assess hypertension."[222])   Moreover, real-word diagnosticity for individuals raises many more issues than just test-retest reliability.   Finally, those with implicit biases need not be regarded as incapable of breaking the causal chain from implicit bias to judgment.   Accordingly, we maintain this scientifically conservative approach and recommend against using the IAT for individual juror selection.[223]

Jury diversity. Consider what a White juror wrote to judge Janet Bond Arterton about jury deliberations during a civil rights complaint filed by Black plaintiffs:

> During deliberations, matter-of-fact expressions of bigotry and broad-brush platitudes about "those people" rolled off the tongues of a vocal majority as naturally and unabashedly as if they were discussing the weather. Shocked and sickened, I sat silently, rationalizing to myself that since I did agree with the product, there was nothing to be gained by speaking out against the process (I now regret my inaction). *Had just one African-American been sitting in that room, the content of discussion would have been quite different.* And had the case been more balanced—one that hinged on fine distinction or subtle nuances—a more diverse jury might have made a material difference in the outcome. I pass these thoughts onto you in the hope that the jury system can some day be improved.[224]

This anecdote suggests that a second-best strategy to striking potential jurors with high implicit bias is to increase jury demographic diversity[225] to get a broader distribution of biases, some of which might cancel each other out.   This is akin to an investment portfolio diversification strategy. Moreover, in a more diverse jury, people's willingness to express explicit biases might be muted, and the very existence of diversity might even affect the operation of implicit biases as well.

---

221.   The test-retest reliability between a person's IAT scores at two different times has been found to be 0.50. For further discussion, see Kang & Lane, *supra* note 2, at 477–78.

222.   *See* Anthony G. Greenwald & N. Sriram, *No Measure Is Perfect, but Some Measures Can be Quite Useful: Response to Two Comments on the Brief Implicit Association Test*, 57 EXPERIMENTAL PSYCHOL. 238, 240 (2010).

223.   For legal commentary in agreement, see, e.g., Anna Roberts, *(Re)forming the Jury: Detection and Disinfection of Implicit Juror Bias*, 44 Conn. L. REV. (forthcoming 2012). Roberts suggests using the IAT during orientation as an educational tool for jurors instead.

224.   Janet Bond Arterton, *Unconscious Bias and the Impartial Jury*, 40 CONN. L. REV. 1023, 1033 (2008) (quoting letter from anonymous juror).

225.   For a structural analysis of why juries lack racial diversity, see Samuel R. Sommers, *Determinants and Consequences of Jury Racial Diversity: Empirical Findings, Implications, and Directions for Future Research*, 2 SOC. ISSUES & POL'Y REV. 65, 68–71 (2008).

In support of this approach, Sam Sommers has confirmed that racial diversity in the jury alters deliberations. In a mock jury experiment, he compared the deliberation content of all-White juries versus racially diverse juries.[226] Racially diverse juries processed information in a way that most judges and lawyers would consider desirable: They had longer deliberations, greater focus on the actual evidence, greater discussion of missing evidence, fewer inaccurate statements, fewer uncorrected statements, and greater discussion of race-related topics.[227] In addition to these information-based benefits, Sommers found interesting predeliberation effects: Simply by knowing that they would be serving on diverse juries (as compared to all-White ones), White jurors were less likely to believe, at the conclusion of evidence but before deliberations, that the Black defendant was guilty.[228]

Given these benefits,[229] we are skeptical about peremptory challenges, which private parties deploy to decrease racial diversity in precisely those cases in which diversity is likely to matter most.[230] Accordingly, we agree with the recommendation by various commentators, including Judge Mark Bennett, to curtail substantially the use of peremptory challenges.[231]

b. Jury Education About Implicit Bias

In our discussion of judge bias, we recommended that judges become skeptical of their own objectivity and learn about implicit social cognition to become motivated to check against implicit bias. The same principle applies to jurors, who must be educated and instructed to do the same in the course of their jury service. This education should take place early and

---

226. These "diverse" juries featured four White and two Black jurors.

227. Samuel R. Sommers, *On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations*, 90 J. Personality & Soc. Psychol. 597 (2006).

228. *See* Sommers, *supra* note 225, at 87.

229. Other benefits include promoting public confidence in the judicial system. *See id.* at 82–88 (summarizing theoretical and empirical literature).

230. Michael I. Norton, Samuel R. Sommers & Sara Brauner, *Bias in Jury Selection: Justifying Prohibited Peremptory Challenges*, 20 J. Behav. Decision Making 467 (2007); Samuel R. Sommers & Michael I. Norton, *Race and Jury Selection: Psychological Perspectives on the Peremptory Challenge Debate*, 63 Am. Psychol. 527 (2008) (reviewing literature); Samuel R. Sommers & Michael I. Norton, *Race-Based Judgments, Race-Neutral Justifications: Experimental Examination of Peremptory Use and the* Batson *Challenge Procedure*, 31 L. & Hum. Behav. 261 (2007) (finding that race influences the exercise of peremptory challenges in participant populations that include college students, law students, and practicing attorneys, and that participants effectively justified their use of challenges in race neutral terms).

231. *See, e.g.,* Mark W. Bennett, *Unraveling the Gordian Knot of Implicit Bias in Jury Selection: The Problems of Judge-Dominated Voir Dire, the Failed Promise of* Batson, *and Proposed Solutions*, 4 Harv. L & Pol'y Rev. 149, 168–69 (2010) (recommending the tandem solution of increased lawyer participation in voir dire and the banning of peremptory challenges); Anthony Page, Batson*'s Blind-spot: Unconscious Stereotyping and the Peremptory Challenge*, 85 B..U. L. Rev. 155 (2005).

often. For example, Judge Bennett spends approximately twenty-five minutes discussing implicit bias during jury selection.[232]

At the conclusion of jury selection, Judge Bennett asks each potential juror to take a pledge, which covers various matters including a pledge against bias:

> I pledge *** :
> I will not decide this case based on biases. This includes gut feeling, prejudices, stereotypes, personal likes or dislikes, sympathies or generalizations.[233]

He also gives a specific jury instruction on implicit biases before opening statements:

> Do not decide the case based on "implicit biases." As we discussed in jury selection, everyone, including me, has feelings, assumptions, perceptions, fears, and stereotypes, that is, "implicit biases," that we may not be aware of. These hidden thoughts can impact what we see and hear, how we remember what we see and hear, and how we make important decisions. Because you are making very important decisions in this case, I strongly encourage you to evaluate the evidence carefully and to resist jumping to conclusions based on personal likes or dislikes, generalizations, gut feelings, prejudices, sympathies, stereotypes, or biases. The law demands that you return a just verdict, based solely on the evidence, your individual evaluation of that evidence, your reason and common sense, and these instructions. Our system of justice is counting on you to

---

232. Judge Bennett starts with a clip from *What Would You Do?*, an ABC show that uses hidden cameras to capture bystanders' reactions to a variety of staged situations. This episode—a brilliant demonstration of bias—opens with a bike chained to a pole near a popular bike trail on a sunny afternoon. First, a young white man, dressed in jeans, a t-shirt, and a baseball cap, approaches the bike with a hammer and saw and begins working on the chain (and even gets to the point of pulling out an industrial-strength bolt cutter). Many people pass by without saying anything; one asks him if he lost the key to his bike lock. Although many others show concern, they don't interfere. After those passersby clear, the show stages its next scenario: a young black man, dressed the same way, approaches the bike with the same tools and attempts to break the chain. Within seconds, people confront him, wanting to know whether the bike is his. Quickly, a crowd congregates, with people shouting at him that he can't take what doesn't belong to him, and some even calling the police. Finally, after the crowd moves on, the show stages its last scenario: a young white woman, attractive and scantily clad, approaches the bike with the same tools and attempts to saw through the chain. Several men ride up and ask if they can help her break the lock! Potential jurors immediately see how implicit biases can affect what they see and hear. *What Would You Do* (ABC television broadcast May 7, 2010), *available at* 2http://www.youtube.com/watch?v=ge7i60GuNRg.

233. Mark W. Bennett, Jury Pledge Against Implicit Bias. In addition, Judge Bennett has a framed poster prominently displayed in the jury room that repeats the language in the pledge.

render a fair decision based on the evidence, not on bias-
es.[234]

Juror research suggests that jurors respond differently to instructions
depending on the persuasiveness of their rationale. For example, jurors
seem to comply more with an instruction to ignore inadmissible evidence
when the *reason* for inadmissibility is potential unreliability, not procedural
irregularity.[235] Accordingly, the implicit bias instructions to jurors should
be couched in accurate, evidence-based, and scientific terms. As with the
judges, the education and instruction should not put jurors on the defen-
sive, which might make them less receptive. Notice how Judge Bennett's
instruction emphasizes the near universality of implicit biases, including in
the judge himself, which decreases the likelihood of insult, resentment, or
backlash from the jurors.

To date, no empirical investigation has tested a system like Judge
Bennett's—although we believe there are good reasons to hypothesize
about its benefits. For instance, Regina Schuller, Veronica Kazoleas, and
Kerry Kawakami demonstrated that a particular type of "reflective" voir
dire, which required individuals to answer an open ended question about
the possibility of racial bias, appeared successful at removing juror racial
bias in assessments of guilt.[236] That said, no experiment has yet been done

---

234. Mark W. Bennett, Jury Instructions Against Implicit Bias. In all criminal cases,
Judge Bennett also instructs on explicit biases using an instruction that is borrowed from a statutory
requirement in federal death penalty cases:

> You must follow certain rules while conducting your deliberations and return-
> ing your verdict:
>
> * * *
>
> Reach your verdict without discrimination. In reaching your verdict, you must
> not consider the defendant's race, color, religious beliefs, national origin, or
> sex. You are not to return a verdict for or against the defendant unless you
> would return the same verdict without regard to his race, color, religious be-
> liefs, national origin, or sex. To emphasize the importance of this requirement,
> the verdict form contains a certification statement. Each of you should careful-
> ly read that statement, then sign your name in the appropriate place in the sig-
> nature block, if the statement accurately reflects how you reached your verdict.

The certification statement, contained in a final section labeled "Certification" on the Ver-
dict Form, states the following:

> By signing below, each juror certifies that consideration of the race, color, reli-
> gious beliefs, national origin, or sex of the defendant was not involved in reach-
> ing his or her individual decision, and that the individual juror would have re-
> turned the same verdict for or against the defendant on the charged offense re-
> gardless of the race, color, religious beliefs, national origin, or sex of the defend-
> ant.

This certification is also shown to all potential jurors in jury selection, and each is asked if
they will be able to sign it.

235. *See, e.g.,* Saul M. Kassin & Samuel R. Sommers, *Inadmissible Testimony, Instructions
to Disregard, and the Jury: Substantive Versus Procedural Considerations*, 23 PERSONALITY & SOC.
PSYCHOL. BULL. 1046 (1997) (finding evidence that mock jurors responded differently to wiretap
evidence that was ruled inadmissible either because it was illegally obtained or not reliable).

236. *See* Regina A. Schuller, Veronica Kazoleas, & Kerry Kawakami, *The Impact of
Prejudice Screening Procedures on Racial Bias in the Courtroom*, 33 L. & HUM. BEHAV. 320 (2009).

on whether jury instructions specifically targeted at implicit bias are effective in real-world settings. Research on this specific question is in development.

  We also recognize the possibility that such instructions could lead to juror complacency or moral credentialing, in which jurors believe themselves to be properly "immunized" or "educated" about bias and thus think themselves to be more objective than they really are. And, as we have learned, believing oneself to be objective is a prime threat to objectivity. Despite these limitations, we believe that implicit bias education and instruction of the jury is likely to do more good than harm, though we look forward to further research that can help us assess this hypothesis.

c.   Encourage Category-Conscious Strategies

  Foreground social categories. Many jurors reasonably believe that in order to be "fair," they should be as colorblind (or gender-blind, etc.) as possible. In other words, they should try to avoid seeing race, thinking about race, or talking about race whenever possible. But the juror research by Sam Sommers demonstrated that White jurors showed race bias in adjudicating the merits of a battery case (between White and Black people) unless the case was made somehow "racially charged." In other words, until and unless White jurors felt there was a specific threat to racial fairness, they showed racial bias.[237]

  What this seems to suggest is that whenever a social category bias might be at issue, judges should recommend that jurors feel free to expressly raise and foreground any such biases in their discussions. Instead of thinking it appropriate to repress race, gender, or sexual orientation as irrelevant to understanding the case, jurors should be made comfortable with the legitimacy of raising such issues. This may produce greater confrontation among the jurors within deliberation, and evidence suggests that it is precisely this greater degree of discussion, and even confrontation, that can potentially decrease the amount of biased decisionmaking.[238]

  This recommendation—to be conscious of race, gender, and other social categories—may seem to contradict some of the jury instructions that we noted above approvingly.[239] But a command that the race (and other social categories) of the defendant should not influence the juror's verdict is entirely consistent with instructions to recognize explicitly that race can

---

237.   *See supra* nn. 68-69.

238.   *See* Alexander M. Czopp, Margo J. Monteith & Aimee Y. Mark, *Standing Up for a Change: Reducing Bias Through Interpersonal Confrontation*, 90 J. Personality & Soc. Psychol. 784, 791 (2006).

239.   *See supra* n. 234 ("[Y]ou must not consider the defendant's race, color, religious beliefs, national origin, or sex. You are not to return a verdict for or against the defendant unless you would return the same verdict without regard to his race, color, religious beliefs, national origin, or sex.") .

have just this impact—unless countermeasures are taken. In other words, in order to make jurors behave in a "colorblind" manner, we can explicitly foreground the possibility of racial bias.[240]

Engage in perspective-shifting. Another strategy is to recommend that jurors try shifting perspectives into the position of the outgroup party, either plaintiff or defendant. Andrew Todd, Galen Bohenhausen, Jennifer Richardson, and Adam Galinsky have recently demonstrated that actively contemplating others' psychological experiences weakens the automatic expression of racial biases.[241] In a series of experiments, the researchers used various interventions to make participants engage in more perspective taking. For instance, in one experiment, before seeing a five-minute video of a Black man being treated worse than an identically situated White man, participants were asked to imagine "what they might be thinking, feeling, and experiencing if they were Glen [the Black man], looking at the world through his eyes and walking in his shoes as he goes through the various activities depicted in the documentary."[242] By contrast, the control group was told to remain objective and emotionally detached. In other variations, perspective-taking was triggered by requiring participants to write an essay imagining a day in the life of a young Black male.

These perspective-taking interventions substantially decreased implicit bias, in the form of negative attitudes, as measured by both a variant of the standard IAT (the personalized IAT) and the standard race attitude IAT.[243] More important, these changes in implicit bias, as measured by reaction time instruments, also correlated with behavioral changes. For example, the researchers found that those in the perspective-taking condition chose to sit closer to a Black interviewer,[244] and physical closeness has long been understood as positive body language, which is reciprocated. Moreover, Black experimenters rated their interaction with White participants put in the perspective-taking condition more positively.[245]

CONCLUSION

Most of us would like to be free of biases, attitudes and stereotypes that lead us to judge individuals based on the social categories they belong

---

240. Although said in a different context, Justice Blackmun's insight seems appropriate here: "In order to get beyond racism we must first take account of race." *Regents of Univ. of Cal. v. Bakke*, 438 U.S. 265, 407 (1978) (Blackmun, J., concurring in part and dissenting in part).

241. Andrew R. Todd et al., *Perspective Taking Combats Automatic Expressions of Racial Bias*, 100 J. PERSONALITY & SOC. PSYCHOL. 1027 (2011).

242. *See id.* at 1030.

243. Experiment one involved the five-minute video. Those in the perspective shifting condition showed a bias of $M=0.43$, whereas those in the control showed a bias of $M=0.80$. Experiment two involved the essay, in which participants in the perspective taking condition showed $M=0.01$ versus $M=0.49$. *See id.* at 1031. Experiment three used the standard IAT. *See id.* at 1033.

244. *See id.* at 1035.

245. *See id.* at 1037.

to, such as race and gender. But wishing things doesn't make them so. And the best scientific evidence suggests that we—all of us, no matter how hard we try to be 'fair and square,' no matter how deeply we believe in our own objectivity—have implicit mental associations that will, in some circumstances, alter our behavior. They manifest everywhere, even in the hallowed courtroom. Indeed, one of our key points here is not to single out the courtroom as a place where bias especially reigns, but rather to suggest that there is no evidence for 'courtroom exceptionalism.' There is simply no legitimate basis for believing that these pervasive implicit biases somehow stop operating in the halls of justice.

Confronted with a robust research basis suggesting the pervasive and widespread effects of bias on decisionmaking, we are therefore forced to choose. Should we seek to be behaviorally realistic, recognize our all-too-human frailties, and design procedures and systems to decrease the impact of bias in the courtroom? Or should we ignore inconvenient facts, stick our heads in the sand, and hope they somehow go away? Even with imperfect information and tentative understandings, we choose the first option. We recognize that our suggestions are starting points, and that they may not all work, and that even as a whole, they may not be sufficient. But we do think they are worth a try. We hope that judges and their institutions agree.