

DOES UNCONSCIOUS RACIAL BIAS AFFECT TRIAL JUDGES?

Jeffrey J. Rachlinski,* Sheri Lynn Johnson,† Andrew J. Wistrich,‡ &
Chris Guthrie††

Race matters in the criminal justice system. Black defendants appear to fare worse than similarly situated white defendants. Why? Implicit bias is one possibility. Researchers, using a well-known measure called the Implicit Association Test, have found that most white Americans harbor implicit bias toward black Americans. Do judges, who are professionally committed to egalitarian norms, hold these same implicit biases? And if so, do these biases account for racially disparate outcomes in the criminal justice system? We explored these two research questions in a multi-part study involving a large sample of trial judges drawn from around the country. Our results—which are both discouraging and encouraging—raise profound issues for courts and society. We find that judges harbor the same kinds of implicit biases as others; that these biases can influence their judgment; but that given sufficient motivation, judges can compensate for the influence of these biases.

© 2009 Jeffrey J. Rachlinski, Sheri Lynn Johnson, Andrew J. Wistrich, and Chris Guthrie. Individuals and nonprofit institutions may reproduce and distribute copies of this Article in any format, at or below cost, for educational purposes, so long as each copy identifies the author, provides a citation to the *Notre Dame Law Review*, and includes this provision and copyright notice.

* Professor of Law, Cornell Law School.

† Professor of Law, Cornell Law School.

‡ Magistrate Judge, United States District Court, Central District of California.

†† Professor of Law, Vanderbilt Law School. The authors are grateful for the comments and assistance of Ian Ayres, Steve Burbank, Jack Glaser, Tracey George, Tony Greenwald, Matthew Patrick Henry, Reid Hastie, Christine Jolls, Dan Kahan, Jerry Kang, Cass Sunstein, and the participants in workshops at the University of Arizona Law School, Bar Ilan University Law School, Brooklyn Law School, the University of Chicago Law School, Chicago-Kent Law School, Cornell Law School, George Washington University Law School, Harvard Law School, Hebrew University Law School, the University of Illinois School of Law, Notre Dame Law School, Ohio State University Law School, St. Louis University Law School, Syracuse University Law School, Tel-Aviv University Law School, Temple Law School, Villanova Law School, the University of Zurich, the Annual Meeting of the American Law and Economics Association, and the Annual Conference on Empirical Legal Studies.

INTRODUCTION

Justice is not blind.

Researchers have found that black defendants fare worse in court than do their white counterparts. In a study of bail-setting in Connecticut, for example, Ian Ayres and Joel Waldfogel found that judges set bail at amounts that were twenty-five percent higher for black defendants than for similarly situated white defendants.¹ In an analysis of judicial decisionmaking under the Sentencing Reform Act of 1984, David Mustard found that federal judges imposed sentences on black Americans that were twelve percent longer than those imposed on comparable white defendants.² Finally, research on capital punishment shows that “killers of White victims are more likely to be sentenced to death than are killers of Black victims” and that “Black defendants are more likely than White defendants” to receive the death penalty.³

Understanding why racial disparities like these and others persist in the criminal justice system is vital. Only if we understand why black defendants fare less well than similarly situated white defendants can we determine how to address this deeply troubling problem.

Two potential sources of disparate treatment in court are *explicit* bias and *implicit* bias.⁴ By explicit bias, we mean the kinds of bias that people knowingly—sometimes openly—embrace. Explicit bias exists and undoubtedly accounts for many of the racial disparities in the criminal justice system, but it is unlikely to be the sole culprit. Researchers have found a marked decline in explicit bias over time, even as disparities in outcomes persist.⁵

Implicit bias—by which we mean stereotypical associations so subtle that people who hold them might not even be aware of them—also appears to be an important source of racial disparities in the criminal

1 Ian Ayres & Joel Waldfogel, *A Market Test for Race Discrimination in Bail Setting*, 46 STAN. L. REV. 987, 992 (1994). To calculate this disparity, Ayres and Waldfogel controlled for eleven other variables, but they conceded that they might still be missing one or more omitted variables that might explain the differential. *Id.* By comparing differences in both bond rates and bail rates, however, they were able to provide even more compelling evidence that the bail rate differences they observed were race-based. *Id.* at 993.

2 David B. Mustard, *Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts*, 44 J.L. & ECON. 285, 300 (2001).

3 R. Richard Banks et al., *Discrimination and Implicit Bias in a Racially Unequal Society*, 94 CAL. L. REV. 1169, 1175 (2006).

4 See Christine Jolls & Cass R. Sunstein, *The Law of Implicit Bias*, 94 CAL. L. REV. 969, 969–70 (2006) (providing examples of both explicit and implicit bias).

5 See PAUL M. SNIDERMAN & THOMAS PIAZZA, *BLACK PRIDE AND BLACK PREJUDICE* 6–8 (2002).

justice system.⁶ Researchers have found that most people, even those who embrace nondiscrimination norms, hold implicit biases that might lead them to treat black Americans in discriminatory ways.⁷ If implicit bias is as common among judges as it is among the rest of the population, it might even account for more of the racially disparate outcomes in the criminal justice system than explicit bias.

In this Article, we report the results of the first study of implicit racial bias among judges. We set out to explore whether judges hold implicit biases to the same extent the general population and to determine whether those biases correlate with their decisionmaking in court. Our results are both alarming and heartening:

- (1) Judges hold implicit racial biases.
- (2) These biases can influence their judgment.
- (3) Judges can, at least in some instances, compensate for their implicit biases.

Our Article proceeds as follows. We begin, in Part I, by introducing the research on implicit bias and its impact on behavior. In Part II, we briefly describe the methods of our study. We provide a much more detailed account in the Appendix. In Part III, we report our results and interpret them. Finally, in Part IV, we explore the implications of our results for the criminal justice system, identifying several possible measures for combating implicit racial bias.

I. IMPLICIT BIAS

Psychologists have proposed that implicit biases might be responsible for many of the continuing racial disparities in society.⁸ To assess the extent to which implicit biases account for racial disparities, researchers must first ascertain whether people hold implicit biases and then determine the extent to which implicit biases influence their actions.

6 Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CAL. L. REV. 945, 951, 961 (2006) (“[E]vidence that implicit attitudes produce discriminatory behavior is already substantial and will continue to accumulate.” (footnote omitted)); Kirstin A. Lane et al., *Implicit Social Cognition and Law*, 3 ANN. REV. L. & SOC. SCI. 427, 433 (2007) (calling implicit social cognitions “robust” and “pervasive”).

7 See Jerry Kang & Mahzarin R. Banaji, *Fair Measures: A Behavioral Realist Revision of “Affirmative Action,”* 94 CAL. L. REV. 1063, 1065 (2006) (arguing that implicit bias shows that affirmative action programs are necessary to address “discrimination in the here and now” (emphasis omitted)).

8 Jerry Kang, *Trojan Horses of Race*, 118 HARV. L. REV. 1489, 1512 (2005).

A. *Demonstrating Implicit Bias*

In their efforts to assess whether people harbor implicit biases, psychologists have used a variety of methods.⁹ Standing front and center among these methods, however, is the Implicit Association Test (IAT).¹⁰ Developed by a research group led largely by Tony Greenwald, Mahzarin Banaji, and Brian Nosek, the IAT is the product of decades of research on the study of bias and stereotypes¹¹ and has attracted enormous scholarly and popular attention.¹² More than four and a half million people have taken the IAT.¹³ The test takes different forms, but most commonly, it consists of a computer-based sorting task in which study participants pair words and faces. A typical administration of the “Race IAT” proceeds as follows¹⁴:

First, researchers present participants with a computer screen that has the words “White or Good” in the upper left-hand corner of the screen and “Black or Bad” in the upper right. The researchers then inform the participants that one of four types of stimuli will appear in the center of the screen: white people’s faces, black people’s faces, good (positive) words, or bad (negative) words. The researchers then explain that the participants should press a designated key on the left side of the computer when a white face or a good word appears and press a designated key on the right side of the computer when a black face or a bad word appears. Researchers refer to the white/good and black/bad pairings as “stereotype congruent,”

9 In addition to the Implicit Association Test, which we discuss in detail, researchers have used subliminal priming techniques, *see, e.g.*, Sandra Graham & Brian S. Lowery, *Priming Unconscious Racial Stereotypes About Adolescent Offenders*, 28 L. & HUM. BEHAV. 483, 487–88 (2004); reaction-time studies, *see, e.g.*, Greenwald & Krieger, *supra* note 6, at 950–53 (labeling studies of implicit bias as studies of biases in reaction times); and novel brain-imaging techniques, *see, e.g.*, Elizabeth A. Phelps et al., *Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation*, 12 J. COGNITIVE NEUROSCI. 729, 729–30 (2000).

10 Alexander R. Green et al., *Implicit Bias Among Physicians and Its Prediction of Thrombolysis Decisions for Black and White Patients*, 22 J. GEN. INTERNAL MED. 1231, 1231–32 (2007).

11 *See* Greenwald & Krieger, *supra* note 6, at 952.

12 *See, e.g.*, Michael Orey, *White Men Can’t Help It*, BUS. WK., May 15, 2006, at 54 (discussing the role of expert witness testimony on “unconscious bias theory” in gender and race employment discrimination cases); Diane Cole, *Don’t Race to Judgment*, U.S. NEWS & WORLD REP., Dec. 26, 2005/Jan. 2, 2006, at 90.

13 *See* Project Implicit, General Information, <http://www.projectimplicit.net/generalinfo.php> (last visited Mar. 9, 2009) (“Visitors have completed more than 4.5 million demonstration tests since 1998, currently averaging over 15,000 tests completed each week.”).

14 Greenwald & Krieger, *supra* note 6, at 952–53 (describing the basic IAT technique).

because they are consistent with negative stereotypes associated with black Americans.¹⁵ The participants complete several trials of this first task.

Then, the computer is programmed to switch the spatial location of “good” and “bad” so that the words “White or Bad” appear in the upper left-hand corner and “Black or Good” appear in the upper right. The researchers explain to the participants that they are now supposed to press a designated key on the left side of the keyboard when a white face or a bad word appears and press a designated key on the right side of the keyboard when a black face or a good word appears. Researchers refer to these white/bad and black/good pairings as “stereotype-*incongruent*,” because they are inconsistent with the negative stereotypes associated with black Americans. The participants then complete several trials of this second task.¹⁶

Researchers have consistently found that white Americans express a strong “white preference” on the IAT.¹⁷ They make this determination by comparing the amount of time it takes respondents to complete the two tasks identified above—that is, their “response latency.”¹⁸ Most white Americans complete the first task (in which they sort white and good from black and bad) more quickly than the second (in which they sort black and good from white and bad).¹⁹ In other words, most white Americans produce higher response latencies when faced with the stereotype-*incongruent* pairing (white/bad or black/good) than when faced with the stereotype-*congruent* pairing (white/good or black/bad).

Researchers have observed a different pattern of implicit biases among black Americans. Black Americans do not exhibit the same white preference that whites express, but neither do they show a mirror-image black preference.²⁰ Rather, black Americans express a much greater variation, with many expressing moderate to strong black preferences that are rarely found in white Americans.²¹ But

15 See Online Psychology Laboratory, Implicit Association Test (Race), <http://opl.apa.org/Experiments/About/AboutIATRace.aspx> (last visited Mar. 9, 2009).

16 See *id.*

17 See Brian A. Nosek et al., *Harvesting Implicit Group Attitudes and Beliefs from a Demonstration Web Site*, 6 GROUP DYNAMICS 101, 105 (2002) (reporting data indicating that white adults taking the IAT strongly favored the white/good versus the black/bad pairing on the IAT).

18 *Id.* at 104.

19 *Id.* at 105.

20 *Id.*

21 *Id.* Throughout, we adopt the convention that a “strong” bias means a tendency to favor one pairing over another on the IAT by over three-quarters of a standard deviation, a “small” bias means an effect of less than one-quarter of a standard

some also express white preferences—sometimes even strong ones.²² On average, black Americans express a slight white preference, but the average masks wide variation in response.²³ Latinos also express a small white preference. Asian Americans show a white preference that is comparable to but somewhat weaker than that found in white Americans.²⁴

The implications of the research using the IAT are a matter of some debate,²⁵ but the cognitive mechanisms underlying the research are clear enough. The white preference arises from well-established mnemonic links. Whites more closely associate white faces with positive words and black faces with negative words than the opposite. Thus, when they complete the white/good versus black/bad trials, they need only make a judgment about whether the stimulus that appears in the middle of the screen is positive or negative. The incongruent association, in contrast, requires that they first judge whether the stimulus is a word or a face and then decide on which side it belongs. Stereotype-incongruent associations interfere with the sorting task in much the same way that the use of green ink can make the word “blue” hard to read.²⁶

The white preference on the IAT is well-documented among white Americans.²⁷ Researchers have conducted and published hundreds of academic studies, and several million people have participated in IAT research.²⁸ They have determined that the implicit biases documented through IAT research are not the product of the order in which people undertake the tasks, their handedness, or any

deviation, and a “moderate” effect means an effect that is in between one-quarter and three-quarters of a standard deviation.

22 *Id.*

23 *Id.*

24 *Id.* at 110.

25 See Hal R. Arkes & Philip E. Tetlock, *Attributions of Implicit Prejudice, or “Would Jesse Jackson ‘Fail’ the Implicit Association Test?”* 15 *PSYCHOL. INQUIRY* 257, 257–58 (2004) (arguing that the IAT does not measure bias or prejudice); Mahzarin R. Banaji et al., *No Place for Nostalgia in Science: A Response to Arkes and Tetlock*, 15 *PSYCHOL. INQUIRY* 279, 279 (2004) (responding to the arguments of Arkes and Tetlock).

26 See J. Ridley Stroop, *Studies of Interference in Serial Verbal Reactions*, 18 *J. EXPERIMENTAL PSYCHOL.* 643, 659–60 (1935) (presenting evidence that words colored differently from their semantic meaning are difficult to read).

27 See Project Implicit, *supra* note 13.

28 *Id.*

other artifact of the experimental method.²⁹ The prevailing wisdom is that IAT scores reveal implicit or unconscious bias.³⁰

B. *Implicit Bias and Behavior*

Even if implicit bias is as widespread as the IAT studies suggest, it does not necessarily lead to, or explain, racially disparate treatment. Only if researchers can show that implicit bias influences decisionmakers can we infer that implicit bias is a cause of racial disparities.

Implicit bias, at least as measured by the IAT, appears to correlate with behavior in some settings. In a recent review, Greenwald and his colleagues identified 122 research reports assessing the relationship between IAT scores and observable behaviors;³¹ of these, thirty-two involved “White-Black interracial behavior.”³² Across these twenty-four studies, the researchers found a modest correlation of 0.24 between the implicit bias measures and the observed behaviors tested in the studies.³³ This means that implicit bias accounted for roughly six percent of the variation in actual behavior.³⁴

Six percent might not sound like much, but a six percent disparity could have an enormous impact on outcomes in the criminal justice system. In a typical year, judges preside over approximately twenty-one million criminal cases in state courts³⁵ and seventy thou-

29 See Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm*, 85 J. PERSONALITY & SOC. PSYCHOL. 197, 209–11 (2003) (discussing mechanisms for reducing order effects); see also Anthony G. Greenwald & Brian A. Nosek, *Health of the Implicit Association Test at Age 3*, 48 ZEITSCHRIFT FÜR EXPERIMENTELLE PSYCHOLOGIE 85, 87 (2001) (“Subject handedness was found to have essentially zero relation to magnitude of the race IAT effect.”).

30 See, e.g., Samuel R. Bagenstos, *Implicit Bias, “Science,” and Antidiscrimination Law*, 1 HARV. L. & POL’Y REV. 477, 477 (2007); Greenwald et al., *supra* note 29, at 199–200.

31 Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity*, J. PERSONALITY & SOC. PSYCHOL. (forthcoming 2009).

32 Note that some of the papers Greenwald and his co-authors include in their analysis report multiple studies using independent samples of subjects. *Id.* (manuscript at 10, 21).

33 *Id.* (manuscript at 21).

34 To be precise, the square of the correlation coefficient of 0.24 is 0.0576, which we round up to 6%.

35 See NAT’L CTR. FOR STATE COURTS, EXAMINING THE WORK OF STATE COURTS, 2006, at 45–46 (Robert C. LaFountain et al. eds., 2006), http://www.ncsconline.org/D_Research/csp/2006_files/EWSC-2007WholeDocument.pdf (providing data for criminal cases entering state courts in 2005).

sand in federal courts,³⁶ many of which involve black defendants. Throughout the processing of these cases, judges make many judgments concerning bail, pretrial motions, evidentiary issues, witness credibility, and so forth. Each of these judgments could be influenced by implicit biases, so the cumulative effect on bottom-line statistics like incarceration rates and sentence length is much larger than one might imagine.³⁷ Furthermore, six percent is only an average. Some judges likely hold extremely strong implicit biases. And some defendants are apt to trigger an unconscious bias to a much greater extent than others.³⁸ Even this seemingly small effect might harm tens or even hundreds of thousands of black defendants every year.

Researchers have found, however, that people may have the ability to compensate for the effects of implicit bias.³⁹ If they are internally driven or otherwise motivated to suppress their own biases, people can make judgments free from biases,⁴⁰ even implicit ones.⁴¹ In one recent study,⁴² for example, a team of researchers administered the IAT to a group of physicians and asked them to diagnose and treat a hypothetical patient—identified to some of the physicians as a white man and to others as a black man—based on a description

36 ADMIN. OFF. OF THE U.S. COURTS, FEDERAL JUDICIAL CASELOAD STATISTICS: MARCH 31, 2007, at 58 tbl.D (2007), <http://www.uscourts.gov/caseload2007/tables/D00CMar07.pdf> (observing U.S. district courts to have 71,652 and 69,697 cases pending in the twelve-month periods ending March 31, 2006 and 2007, respectively).

37 Kang & Banaji, *supra* note 7, at 1073.

38 See Jennifer Eberhardt et al., *Looking Deathworthy: Perceived Stereotypicality of Black Defendants Predicts Capital-Sentencing Outcomes*, 17 PSYCHOL. SCI. 383, 384 (2006) (“Defendants whose appearance was perceived as more stereotypically Black were more likely to receive a death sentence than defendants whose appearance was perceived as less stereotypically Black.”).

39 See Jack Glaser & Eric D. Knowles, *Implicit Motivation to Control Prejudice*, 44 J. EXPERIMENTAL SOC. PSYCHOL. 164, 164–65, 170–71 (2008).

40 See Bridget C. Dunton & Russell H. Fazio, *An Individual Difference Measure of Motivation to Control Prejudiced Reactions*, 23 PERSONALITY & SOC. PSYCHOL. BULL. 316, 324–26 (1997); E. Ashby Plant & Patricia G. Devine, *Internal and External Motivation to Respond Without Prejudice*, 75 J. PERSONALITY & SOC. PSYCHOL. 811, 824–28 (1998).

41 See John A. Bargh, *The Cognitive Monster: The Case Against the Controllability of Automatic Stereotype Effects*, in DUAL-PROCESS THEORIES IN SOCIAL PSYCHOLOGY 361, 375–78 (Shelly Chaiken & Yaacov Trope eds., 1999); Patricia G. Devine et al., *The Regulation of Explicit and Implicit Race Bias: The Role of Motivations to Respond Without Prejudice*, 82 J. PERSONALITY & SOC. PSYCHOL. 835, 845–47 (2002); John F. Dovidio et al., *On the Nature of Prejudice: Automatic and Controlled Processes*, 33 J. EXPERIMENTAL SOC. PSYCHOL. 510, 535–36 (1997); Russell H. Fazio et al., *Variability in Automatic Activation as an Unobtrusive Measure of Racial Attitudes: A Bona Fide Pipeline?*, 69 J. PERSONALITY & SOC. PSYCHOL. 1013, 1025–26 (1995).

42 Green et al., *supra* note 10.

of symptoms.⁴³ The researchers found a correlation between IAT scores and treatment; the physicians with higher IAT scores were more likely to offer appropriate treatment to white patients than to black patients diagnosed with the same condition.⁴⁴ But among the sixty-seven physicians who reported some awareness of the purpose of the study, those with higher IAT scores were more likely to recommend the treatment to black patients.⁴⁵ In other words, the doctors who were aware of the purpose of the study compensated for their implicit biases when the situation made them sensitive to the risk of behaving—or being observed to behave—in a biased way. “This suggests,” argue the authors, “that implicit bias can be recognized and modulated to counteract its effect on treatment decisions.”⁴⁶

Jack Glaser and Eric Knowles found similar results in a study using the so-called “Shooter Task.”⁴⁷ In research of this type, subjects participate in a simulation akin to a video game in which they watch a person on screen pull either a gun or an innocent object, like a wallet, out of his pocket.⁴⁸ If he pulls a gun, the participants are instructed

43 *Id.* at 1232–33.

44 *Id.* at 1235. The researchers also found that white doctors who express white preferences on the IAT were *more* likely to diagnose black patients than white patients as having coronary artery disease, based upon the same symptoms. *Id.* at 1234–35. Indeed, the doctors offered the appropriate treatment—thrombolysis—to an equal number of black patients as white patients! *Id.* As the authors rightly point out, this does not mean there was no disparity; among patients who were diagnosed as suffering from coronary artery disease, black patients were less likely to be offered the appropriate treatment. *Id.* It is at least curious, however, that doctors with implicit white preferences would be more likely to diagnose coronary artery disease for black patients than white patients, but less likely to treat it. The diagnosis disparity runs in the opposite direction of the treatment-for-diagnosis disparity, and ultimately, the two effects actually cancel each other out. *Id.* at 1236–37. Of course, if doctors behaved the same way in the real world, *black and white patients who presented the same symptoms would be treated in the same way.* Thus, though the IAT predicted discriminatory acts, implicit bias does not seem to result in discrimination overall. *Id.* at 1234–37. This aspect of the study has been the source of some debate. See John Tierney, *In Bias Test, Shades of Gray*, N.Y. TIMES, Nov. 18, 2008, at D1. One other recent study also shows no correlation between measures of implicit bias and medical decisions among physicians. See Janice A. Sabin et al., *Physician Implicit Attitudes and Stereotypes About Race and Quality of Medical Care*, 46 MED. CARE 678, 682 (2008) (“We did not find a relationship between difference in treatment recommendations by patient race and implicit measures.”).

45 Green et al., *supra* note 10, at 1235.

46 *Id.* at 1237.

47 Glaser & Knowles, *supra* note 39, at 167–71.

48 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).

to “shoot” by pushing a button on a joystick; if he pulls a benign object, they are instructed to refrain from shooting.⁴⁹ Researchers have found that most white adults exhibit a “shooter bias” in that they are more likely to shoot a black target—regardless of what object the on-screen target pulls out of his pocket⁵⁰—and that this effect correlates with a white preference on the IAT.⁵¹ Glaser and Knowles found in their study, however, that those rare individuals with a white preference on the IAT and who are highly motivated to control prejudice were able to avoid the shooter bias.⁵² In short, “those high in an implicit negative attitude toward prejudice show less influence of implicit stereotypes on automatic discrimination.”⁵³

In sum, the research on implicit bias suggests that people exhibit implicit biases, that there is some evidence that implicit bias can influence behavior, and that people can overcome or compensate for implicit biases if properly motivated and if the racial context is made sufficiently salient. Whether and how this research applies to judges and the criminal justice system is an open question and one to which we turn in the next Part.

II. THE STUDY DESIGN

We are aware of only two IAT studies exploring a behavior of direct interest to the criminal justice system. In one study, researchers found that college student subjects harboring a strong implicit bias in favor of whites imposed longer criminal sentences on a Latino defendants than on a white defendants.⁵⁴ In another study in Germany, researchers correlated implicit attitudes towards native Germans and Turkish immigrants among German college students with judgments of guilt of a Turkish defendant.⁵⁵ The researchers found a high correlation between negative association with Turkish immigrants and judgments of guilt when the materials made “threatening” aspects of the

49 *Id.* at 1315–16.

50 *Id.* at 1320.

51 *Id.* at 1320–21; Glaser & Knowles, *supra* note 39, at 168–69.

52 Glaser & Knowles, *supra* note 39, at 169–70.

53 *Id.* at 171.

54 Robert W. Livingston, When Motivation Isn’t Enough: Evidence of Unintentional Deliberative Discrimination Under Conditions of Response Ambiguity 9–10 (2002) (unpublished manuscript, on file with the *Notre Dame Law Review*).

55 See Arnd Florack et al., *Der Einfluss Wahrgenommener Bedrohung auf die Nutzung Automatischer Assoziationen bei der Personenbeurteilung* [The Impact of Perceived Threat on the Use of Automatic Associations in Person Judgments], 32 ZEITSCHRIFT FÜR SOZIALPSYCHOLOGIE 249 (2001).

Turkish defendant salient.⁵⁶ Though suggestive, these studies, standing alone, do not tell us much about implicit bias in the criminal justice system. Most importantly, they tell us nothing about a central actor in the system: the judge. Do judges hold implicit racial biases? If so, do those biases affect their judgments in court? We sought to answer these two questions in our study.⁵⁷

A. Judges

We recruited judges to participate in our study at judicial education conferences, as we have in our prior work.⁵⁸ The 133 judges who participated in our study came from three different jurisdictions.⁵⁹ The judges asked us not to identify their jurisdictions,⁶⁰ but we can describe the basic characteristics of each of the three. We recruited seventy judges from a large urban center in the eastern United

56 *Id.* at 255 tbl.1.

57 We recognize that we have emphasized disparities concerning black Americans, rather than other races. We have done so for three reasons. First, even though Latinos, Native Americans, and Asian Americans are also targets of racism, both explicit and implicit, in the United States some of the most striking disparities involve black Americans in the legal system. Second, the research on the IAT has emphasized biases concerning black Americans as well. Third, our sample of judges includes a large group of black American judges, but few Latinos, few Asian Americans, and no Native Americans. We thus cannot draw any conclusions about the reactions of judges of these ethnicities. We therefore focus our attention here on biases involving black Americans.

58 See Chris Guthrie et al., *Blinking on the Bench: How Judges Decide Cases*, 93 CORNELL L. REV. 1, 13 (2007) [hereinafter Guthrie et al., *How Judges Decide*]; Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 814–15 (2001) [hereinafter Guthrie et al., *Judicial Mind*]; Jeffrey J. Rachlinski et al., *Inside the Bankruptcy Judge's Mind*, 86 B.U. L. REV. 1227, 1256–59 (2006); Andrew J. Wistrich et al., *Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding*, 153 U. PA. L. REV. 1251, 1323–24 (2005).

59 At two of the conferences, we collected data from judges attending a plenary session. At the third, we collected data from judges attending an optional session.

60 Their concerns might be justified. Some of our previous work has been reported in the *New York Times* and the *American Bar Association Journal*, among other places. See, e.g., Patricia Cohen, *Judicial Reasoning Is All Too Human*, N.Y. TIMES, June 30, 2001, at B7; Debra Cassens Weiss, *Judges Flunk Story Problem Test, Showing Intuitive Decision-Making*, A.B.A. J., Feb. 19, 2008, https://abajournal.com/news/judges_flunk_story_problem_test_showing_intuitive_decision_making/. The latter report leads with the unfortunate headline “Judges Flunk Story Problem Test,” which casts the judges in a more negative light than the data warrant. Interest in the present Article is sufficiently high that, despite our own efforts to limit its use before it was finalized, it was cited by Judge Jack Weinstein in a published opinion, *United States v. Taveras*, 424 F. Supp. 2d 446, 462 (E.D.N.Y. 2006), and discussed at length in a recent volume of the *Annual Review of Law and Social Science*, Lane et al., *supra* note 6, at 441–45.

States.⁶¹ These seventy judges, who are appointed to the bench for renewable terms, constitute roughly three-quarters of the judges who sit in this jurisdiction. We recruited forty-five judges from a large urban center in the western United States.⁶² These forty-five judges, who are appointed to the bench but then stand for election, make up roughly half of the judges in their jurisdiction. We recruited our final group of judges at an optional session at a regional conference. These eighteen judges, who sit in various towns and cities throughout the state in which the conference was held, are appointed to the bench but are then required to stand for election.⁶³

We did not ask the judges to identify themselves by name, but we did ask them to identify their race, gender, exact title, political affiliation, and years of experience on the bench.⁶⁴ Table 1 summarizes the demographic information that the judges provided. As Table 1 indicates, our sample of judges, particularly those from the eastern jurisdiction, is fairly diverse, at least in terms of gender and race.

61 Eighty judges attended the session at which we collected data, but we excluded ten from our study. We excluded one judge at his or her request. We excluded nine other judges because they failed to provide us with demographic information. We believe that these failures were largely accidental. To complete the demographic page, the judges had to return to the written materials after completing the final IAT, and these nine judges failed to do so. We did not realize that this process would cause problems at our presentation in the eastern jurisdiction, and hence we did not obtain this data. In the subsequent presentations, we made sure that the judges completed the last page as we collected the surveys.

62 Forty-eight judges attended the session at which we collected the data, but we excluded three from our study. One judge neglected to provide demographic information, and we lost the data for two other judges due to a computer malfunction.

63 Over ninety percent of the judges in the eastern jurisdiction attended this conference (although, as noted, we did not obtain data from all of them). Attendance was lower among the western judges; the sample includes roughly half of the judges in their jurisdiction. These judges' willingness to participate in our study was thus unlikely to have been affected by their interest (or lack thereof) in the content of the material. In fact, the judges were not aware of the subject matter of the talk before the session began. This was not our first presentation to the eastern judges. Three years earlier, we had presented a completely different set of materials to the same educational conference. Some of the results from that earlier session have been published, also without identifying the jurisdiction. Wistrich et al., *supra* note 58, at 1279–81. Many of the judges were therefore familiar with our methods, although the present study differs from our earlier work. Our prior work dealt largely with judicial reliance on heuristics in making judgments, whereas this research is entirely devoted to the influence of race and gender on judgment. This was our first presentation to the western judges. The regional judges differed from the eastern and western judges in that they opted not only to attend the judicial education conference at which we spoke but also to attend our optional session.

64 We include these questions below in Appendix A.

TABLE 1: DEMOGRAPHIC INFORMATION OF THE JUDGES
(PERCENTAGE WITHIN GROUP AND NUMBER)

Demographic Parameter		Eastern Jurisdiction (70)	Western Jurisdiction (45)	Optional Conference (18)	Overall (133)
Race	White	52.9 (37)	80.0 (36)	66.7 (12)	63.9 (85)
	Black	42.9 (30)	4.4 (2)	5.6 (1)	24.8 (33)
	Latino	4.3 (3)	11.1 (5)	16.7 (3)	8.3 (11)
	Asian	0.0 (0)	4.4 (2)	11.1 (2)	3.0 (4)
Gender	Male	55.7 (39)	66.7 (30)	50.0 (9)	58.7 (78)
	Female	44.3 (31)	33.3 (15)	50.0 (9)	41.4 (55)
Political Affiliation	Democrat	86.6 (58)	64.4 (29)	64.7 (11)	76.0 (98)
	Republican	13.4 (9)	35.6 (16)	35.3 (7)	24.0 (31)
Average Years of Experience		9.8	10.8	9.3	10.1

B. Methods and Materials

To explore the two questions animating this Article—that is, whether judges hold implicit racial biases, and if so, whether those biases produce biased judicial decisions—we designed a multipart study requiring the participating judges to complete computer tasks⁶⁵ and then to respond to a paper questionnaire.

We proceeded as follows. We placed in front of each judge a laptop computer and a questionnaire. The computer screen and the front page of the questionnaire introduced the study and asked the judges to await instruction before beginning.⁶⁶ Once the judges were

⁶⁵ The computer tasks were all conducted on laptop computers rented for the purpose of running the experiment. They were all relatively contemporary machines of similar makes. At the eastern and western sessions, all were Hewlett-Packard NX9010; at the regional conference, they were IBM ThinkPads. All had fifteen-inch screens. The software to run the tasks was designed with a program called Inquisit 2.0, created specifically for measuring implicit associations by a company called Millisecond Software. See Inquisit, <http://www.millisecond.com> (last visited Mar. 7, 2009).

⁶⁶ The instructions on the survey were as follows:

Many of the points to be discussed at this session are best experienced directly. We therefore ask that before the session starts, you participate in a series of exercises on the laptop computer and evaluate a series of hypothetical cases in the pages that follow. (Participation in all aspects of this exercise is voluntary, of course.) Please do not discuss these materials while you are participating. We shall collect these surveys before the discussion and present the results during the session.

The first part of the exercise consists of a computer task. Please do not begin the task or turn this page until asked to do so.

The instructions on the computer screen were:

fully assembled, we announced “Today, we shall ask you to participate actively in your own education.”⁶⁷

We asked the judges to complete the computer tasks and to respond to the questionnaire according to the instructions provided. We assured the judges that their responses were anonymous and that we had no way of identifying them individually, but we also made clear that participation was entirely voluntary and that any judge who wanted to exclude her results from the study could do so. (Only one judge chose to do so.) We informed the judges that we would compile their cumulative results and share them with the group at the end of the session.

With these important preliminaries out of the way, we then asked the judges to begin the study. The study included a race IAT;⁶⁸ two hypothetical vignettes in which the race of the defendant was not explicitly identified but was subliminally primed; and another hypothetical vignette in which the race of the defendant was made explicit.⁶⁹ The final page of the questionnaire asked judges to provide the basic demographic information identified above.⁷⁰

III. THE STUDY RESULTS

We present the results in two parts. First, we report the judges’ IAT scores, which demonstrate that judges, like the rest of us, harbor implicit racial biases. Second, we report the results of our judicial decisionmaking studies, which show that implicit biases can influence judicial decisionmaking but can also be overcome, at least in our experimental setting.⁷¹

JURISDICTION: Judicial Education Conference, DATE

We shall begin by making announcements as to the nature of this exercise.

Please DO NOT BEGIN until after the announcements.

After the announcements, please press the space bar to begin.

67 Judge Wistrich conducted the introduction at the eastern and western conferences; Professor Rachlinski did it at the regional conference.

68 We also conducted an IAT related to gender after the race IAT, but do not report those results here.

69 We also included a scenario in which we manipulated the gender of a target legal actor as the third scenario. We do not report these results here.

70 The order of the materials was thus as follows: the priming task; the written scenario of the shoplifter; the written scenario of the armed robber; the gender scenario (not reported here); the battery case; the race IAT; the gender IAT (not reported here); and the demographics page.

71 We analyzed the three groups of judges separately, but there were no significant differences between the judges, except as noted below, so we have kept them together throughout the analysis. Similarly, we found no differences between the judges on the basis of the gender, political affiliation, or experience. Because previ-

A. *The Implicit Association Test*

To measure implicit associations involving race, we gave the judges a computer-based-race IAT comparable to the race IAT given to millions of study participants around the world.⁷² We asked the judges to perform two trials of the IAT, as described above. The first required them to pair white faces with positive words and black faces with negative words. In other words, the first trial required them to select stereotype-congruent pairings. The second required them to pair white faces with negative words and black faces with positive words. In other words, the second trial required them to select stereotype-incongruent pairings.⁷³

To determine each judge's implicit bias score, we performed two calculations. First, we subtracted each judge's average response latency in the stereotype-congruent round from the stereotype-incongruent round to calculate the IAT measure. This measure reflects the most commonly used scoring method for large samples of data collected on the Internet, and hence allows us to compare judges to ordinary adults.⁷⁴ Second, we constructed a standardized measure consisting of the average difference in response latencies for each judge divided by the standard deviation of that judge's response latencies in the target rounds. This measure is less commonly reported, but more stable, and produces higher correlations with other behaviors.⁷⁵

ous research on the IAT suggests that Latinos score somewhat closer to black Americans on the IAT we used, we combined the few Latino judges with the black judges for these analyses. Nosek et al., *supra* note 17, at 110 tbl.2. Similarly, we combined the Asian American judges with the white judges.

72 The exact instructions at the outset of the IAT were as follows:

The remaining computer tasks involve making CATEGORY JUDGMENTS. Once the tasks begin, a word or words describing the CATEGORIES will appear in the upper left and upper right corners of the computer screen. A TARGET word or picture will also be displayed in the center of the screen, which you must assign to one of the two categories. Please respond AS RAPIDLY AS POSSIBLE, but don't respond so fast that you make many errors. (Occasional errors are okay.) An "X" will appear when you make an error. Whenever the "X" appears, correct the mistake by pressing the other key.

73 For a more detailed account of our IAT procedure, see Appendix B.

74 See, e.g., Nosek et al., *supra* note 17, at 104–05 (reporting average differences in response latencies among large samples of subjects obtained through the Internet).

75 See Greenwald et al., *supra* note 29, at 209–10 (describing standardized measures). The full account of our scoring methods is included as Appendix C.

We found a strong white preference among the white judges, as shown in Table 2. Among the eighty-five white judges, seventy-four (or 87.1%) showed a white preference on the IAT. Overall, the white judges performed the stereotype-congruent trial (white/good and black/bad) 216 milliseconds faster than the stereotype-incongruent trial (black/good and white/bad). The black judges, by contrast, demonstrated no clear preference overall. Although fourteen of forty-three (or 44.2%) showed a white preference, the black judges performed the stereotype-congruent trial (white/good and black/bad) a mere twenty-six milliseconds faster than the stereotype-incongruent trial (black/good and white/bad). Comparing the mean IAT scores of the white judges with those of the black judges revealed that the white judges expressed a significantly larger white preference.⁷⁶

TABLE 2: RESULTS OF RACE IAT BY RACE OF JUDGE

<i>Race of Judge (sample size)</i>	<i>Mean IAT Score in milliseconds (and standard deviation)*</i>		<i>Percent of Judges with lower average latencies on the white/good versus black/bad round</i>
	<i>Judges</i>	<i>Internet Sample</i>	
<i>White (85)</i>	216 (201)	158 (224)	87.1
<i>Black (43)</i>	26 (208)	39 (244)	44.2

*Note: Positive numbers indicate lower latencies on the white/good versus black/bad round

Because we used a commonly administered version of the IAT, we are able to compare the results of our study to the results of other studies involving ordinary adults. We found that the black judges produced IAT scores comparable to those observed in the sample of black subjects obtained on the Internet.⁷⁷ The white judges, on the other

⁷⁶ The specific statistical result was: $t(82) = 4.94$, $p < .0001$. Throughout this Article, we reserve the use of the words “significant” and “significantly” for statistical significance.

⁷⁷ The specific statistical result was: $t(42) = 0.18$, $p = .86$. In conducting this test, we took the effect size among the Internet sample of 0.16 standard deviations to be the “population” effect size among black participants on the Internet, and tested whether our observed difference, with our observed standard deviation, would be likely to be reliably higher or lower than the effect in the Internet data. The priming condition did not appear to affect the judges’ IAT scores. Also, the judges themselves varied somewhat in their IAT scores. White judges in the eastern jurisdiction expressed an average standardized preference of 0.33, compared to 0.48 and 0.55 in the western jurisdiction and the regional conferences, respectively. These differences were marginally significant. Because the black judges in our study were concentrated largely in the eastern jurisdiction, similar tests for variations among these judges would not be reliable.

hand, demonstrated a statistically significantly stronger white preference than that observed among a sample of white subjects obtained on the Internet.⁷⁸ For two reasons, however, this does not necessarily mean that the white judges harbor more intense white preferences than the general population. First, we did not vary the order in which we presented the materials, and this order effect could have led to artificially higher IAT scores.⁷⁹ Second, the judges performed both trials much more slowly than the other adults with whom we are making this comparison, and this, too, could have led to artificially higher IAT scores.⁸⁰ We also suspect that the judges were older, on average, than the Internet sample. To the extent that implicit racial bias is less pronounced among younger people, we would expect the judges to exhibit more implicit bias than the Internet sample.

B. IAT and Judicial Behavior

To assess the impact of implicit bias on judicial decisionmaking, we gave the judges three hypothetical cases: the first involving a juvenile shoplifter, the second involving a juvenile robber, and the third involving a battery. We speculated that the judges might respond differently depending upon whether we made the race of the defendant salient, so in the first two cases, we did not identify the race of the defendant explicitly, but we did so implicitly through a subliminal priming technique described below. In the third case, we made race explicit, informing some of the judges that the defendant was “Caucasian” and others that he was “African American.”⁸¹ By comparing the

78 The specific statistical result was: $t(84) = 2.26$, $p = .026$. We compared our results to those of the Internet sample reported in Nosek et al., *supra* note 17, at 105. In making this comparison, we took the effect size among the Internet sample of 0.83 standard deviations to be the “population” effect size among white participants on the Internet, and tested whether our observed difference, with our observed standard deviation, would likely be reliably higher or lower than the effect in the Internet data.

79 We selected data collection and scoring procedures so as to minimize the effects of order of presentation. Greenwald and his fellow authors reported that the effect of order of presentation is less than one percent, using the methods we followed. See Greenwald et al., *supra* note 29, at 210 tbl.2.

80 See *id.* at 200 (“IAT effects will be artificially larger for any subjects who respond slowly.”).

81 Throughout this Article we follow the convention of using the terms “black” and “white” to denote race, as the terms more closely reflect the faces in the IAT, the instructions in the IAT (which refer to black and white), and might more closely reflect how the black judges would describe themselves (although there would be variation on this). When referring to the criminal defendants, however, we use African American and Caucasian, following the references mentioned in the hypothetical cases.

judges' individual IAT scores with their judgments in these hypothetical cases, we are able to assess whether implicit bias correlates with racially disparate outcomes in court.

1. Race Primed

We asked the judges to decide two hypothetical cases, one involving a juvenile shoplifter and one involving a juvenile armed robber. Before giving the judges the scenarios, though, we asked them to perform a subliminal priming task, following a protocol developed by Sandra Graham and Brian Lowery.⁸² The task appeared to be a simple, computer-based, spatial recognition task.⁸³ To complete the task, the judges were required to focus their attention on the center of the computer screen in front of them. Words appeared in one of the four corners for 153 milliseconds before being masked by a string of random letters.⁸⁴ At that speed, words are extremely difficult to process

82 Graham & Lowery, *supra* note 9, at 487–88.

83 At the beginning of the task, three asterisks appeared in the center of the screen. A sixteen-character letter string then appeared in one of the four quadrants of the screen. The judges were instructed to press a specific key on the left-hand side of the computer (the “E” key, which was marked with a red dot) when the letter string appeared in one of the quadrants on the left and to press a specific key on the right-hand side of the computer (the “I” key, which was also marked with a red dot) when a word appeared in one of the two quadrants on the right. Reminders as to which key to press also remained on the computer screen throughout the first task (that is, “press the ‘E’ key for left” and “press the ‘I’ key for right”). When the judges identified the quadrant correctly, the word “correct” would appear in the center in letters. When the judges made an error, the word “error” would appear instead. In either case, the three asterisks would then replace the words “correct” or “error” and the task would repeat. The exact instructions the judges saw are below.

Once you begin the first computer task, the screen will go blank, then three asterisks (* * *) will appear in the center. Focus your attention on these. A string of letters will then appear in the upper-right, lower-right, upper-left, or lower-left portion of the computer screen.

If the string appears on the left-hand side (either up or down), press the “E” key.

If the string appears on the right-hand side (either up or down), press the “I” key.

If you correctly identify the position, the screen will flash the word “correct”; if you identify the wrong position, the screen will flash the word “error.”

The task will then repeat a number of times. Other words may appear with the letter string. Ignore these and try to identify the position of the letters as quickly as possible.

When you are ready, press the space bar to begin the task.

84 Each trial thus proceeded as follows: the three asterisks would appear in the center of the screen; 1200 milliseconds later (1.2 seconds) one of the prime words (selected at random) would appear in one of the four quadrants (at random as deter-

consciously.⁸⁵ Each judge saw sixty words. Half of the judges saw words associated with black Americans,⁸⁶ and half saw words with no common theme.⁸⁷ After the sixtieth trial, the task stopped.⁸⁸ The computer screen then instructed the judges to turn to the written materials.⁸⁹

mined by the computer); 153 milliseconds after that, the letter-string would appear over the prime; this would remain until the judge pressed either the “E” or “I” key; then either the “correct” or “error” in the center (depending upon the judge’s response) and would remain for roughly one second; then the three asterisks would replace the word “correct” or “error”; and the process would repeat. Due to an error in the computer programming, the judges in the eastern conference were only exposed to the subliminal prime for sixty-four milliseconds, rather than 153 milliseconds.

85 Graham and Lowery reported that none of the officers in their study was able to identify the nature of the words being shown to them. Graham & Lowery, *supra* note 9, at 491. We did not ask our judges their assessment of what the words were.

86 The words came directly from the Graham and Lowery study: graffiti, Harlem, homeboy, jerricurl, minority, mulatto, negro, rap, segregation, basketball, black, Cosby, gospel, hood, Jamaica, roots, afro, Oprah, Islam, Haiti, pimp, dreadlocks, plantation, Tyson, welfare, athlete, ghetto, calypso, reggae, rhythm, soul. *Id.* at 489 n.5.

87 These words also came directly from Graham and Lowery: baby, enjoyment, heaven, kindness, summer, sunset, truth, playful, accident, coffin, devil, funeral, horror, mosquito, stress, toothache, warmth, trust, sunrise, rainbow, pleasure, paradise, laughter, birthday, virus, paralysis, loneliness, jealousy, hell, execution, death, agony. Graham and Lowery used neutral words that matched the words associated with black Americans for positive or negative associations. *Id.*

88 Our study differed from that of Graham and Lowery in several ways, any of which might have affected the results. First, Graham and Lowery used eighty trials, rather than the sixty we used. *Id.* at 489–90. Second, because we ran a large group of judges at the same time, we did not use audible beeps to indicate correct responses. *Id.* Third, our hypothetical defendants differed. We did not have access to the original materials Graham and Lowery used, and so wrote our own. See fact pattern *infra* Appendix A. Fourth, we asked fewer questions concerning the hypothetical defendants. Although we do not see how any of these differences would necessarily affect the results, priming tasks can be sensitive to details.

89 The following appeared on the screen:

Thank you for completing the first computer task.

Now please turn to the written materials.

Please leave this computer on with the screen up.

After you have completed four pages of written materials, please press the space bar to continue with the final computer tasks.

In case a judge accidentally or mistakenly hit the space bar, we added another intervening page before the second computer task, which appeared once the space bar was pressed. It read as follows:

If you have completed the four case summaries, please press the space bar to begin the final computer task.

a. The Shoplifter Case

We first presented the judges with a scenario called the “Shoplifter Case.” The judges learned that William, a thirteen year old with no prior criminal record, had been arrested for shoplifting several toys from a large, upscale toy store.⁹⁰ The judges read that there is some conflicting evidence on the degree to which William resisted arrest, but there is no dispute over the fact that he had shoplifted.⁹¹

Following the scenario, we asked the judges three questions about William. First, we asked them what disposition they thought most appropriate. We listed seven options below the question, ranging from a dismissal of the case to a transfer to adult court.⁹² Second, we asked judges to predict on a seven-point scale (from “Not at all Likely” to “Very Likely”) whether William would commit a similar crime in the future. And finally, we asked them to predict on an identical seven-point scale the likelihood that William would commit a more serious crime in the future. In short, we asked them one question about sentencing and two questions about recidivism.

The judges’ determinations were not influenced by race. As shown in Table 3, judges primed with the black-associated words did not produce significantly different judgments than the judges primed with the neutral words.⁹³ Our primary interest, however, was in determining whether the judges’ implicit biases correlated with their judgments. We found that the judges’ scores on the race IAT had a marginally significant influence on how the prime influenced their judgment.⁹⁴ Judges who exhibited a white preference on the IAT gave

90 The location of the crime would reveal the jurisdiction and hence we delete it. The location was an upscale shopping district.

91 The exact materials for this scenario and all others are included *infra* Appendix A.

92 The options were as follows:

- (1) Dismiss it with an oral warning
- (2) Adjourn the case in contemplation of dismissal (assuming William gets in no further trouble)
- (3) Put William on probation for six months or less
- (4) Put William on probation for more than six months
- (5) Commit William to a juvenile detention facility for six months or less
- (6) Commit William to a juvenile detention facility for more than six months
- (7) Transfer William to adult court.

93 The results were as follows: Question 1, $z = 0.51$, $p = .61$; Question 2, $z = 0.73$, $p = .46$; Question 3, $z = 1.09$, $p = .28$.

94 To accomplish this analysis, we conducted an ordered logit regression of the judges’ disposition against the priming condition, the judges’ IAT scores, and an interaction of the two. The interaction term reflects the effect of the IAT score on

TABLE 3: AVERAGE RESULTS ON JUVENILE SHOPLIFTER (ALL THREE QUESTIONS ON A SEVEN-POINT SCALE: HIGHER NUMBERS INDICATE HARSHER JUDGMENTS*)

<i>Prime (and n)</i>	<i>Q1: Disposition</i>	<i>Q2: Recidivism-Same Crime</i>	<i>Q3: Recidivism-More Serious Crime</i>
<i>Black (63)</i>	2.34	2.58	2.23
<i>Neutral (70)</i>	2.40	2.36	1.94

*Note: The seven-point scale for questions two and three have been transposed from the original for this Table, so that higher numbers consistently meant harsher judgment.

harsher sentences to defendants if they had been primed with black-associated words rather than neutral words, while judges who exhibited a black preference on the IAT gave less harsh sentences to defendants if they had been primed with black-associated words rather than neutral words. We did not find any significant relationship between the judges’ IAT scores and either of the recidivism measures, although the data showed a similar trend.⁹⁵

b. The Robbery Case

The second scenario, called the “Robbery Case,” described Michael, who was arrested for armed robbery at a gas station convenience store two days shy of his seventeenth birthday.⁹⁶ Michael, who had previously been arrested for a fight in the school lunchroom, threatened the clerk at the convenience store with a gun and made off with \$267 in cash. He admitted the crime, claiming that his friends had dared him to do it. After they had read this scenario, we asked the judges the same three questions we asked them about William in the shoplifter case.

Again the judges’ determinations were not influenced by race. As shown in Table 4, the judges primed with black-associated words did not produce significantly different ratings than the judges primed

how the prime affected the judge. This term was marginally significant in the model, $z = 1.84$, $p = .07$.

95 For the first recidivism question, $z = 1.41$, $p = .16$. On the second recidivism question, $z = 1.49$, $p = .14$. On these questions, the black judges and the white judges seemed to respond in similar ways. We ran the full model (predictors of prime, race of judge, IAT, and all interactions between these variables) on all three variables as well. Adding the race-of-judge terms and interactions did not produce any significant effects.

96 The use of an armed robbery breaks somewhat with Graham and Lowery, who had used two simple property crimes. See Graham & Lowery, *supra* note 9, at 490.

TABLE 4: AVERAGE RESULTS ON JUVENILE ARMED ROBBER (ALL THREE QUESTIONS ON A SEVEN-POINT SCALE: HIGHER NUMBERS INDICATE HARSHER JUDGMENTS*)

<i>Prime (and n)</i>	<i>Q1: Disposition</i>	<i>Q2: Recidivism-Same Crime</i>	<i>Q3: Recidivism-More Serious Crime</i>
<i>Black (63)</i>	4.92	3.54	3.17
<i>Neutral (70)</i>	4.97	3.61	3.48

*Note: The seven-point scale for questions two and three have been transposed from the original for this Table, so that higher numbers consistently meant harsher judgment.

with the neutral words.⁹⁷ As noted, however, our primary interest was in the relationship between implicit bias and these judgments. As with the shoplifting case, the judges' scores on the race IAT had a marginally significant influence on how the prime influenced their judgment in the robbery case.⁹⁸ Judges who exhibited a white preference on the IAT gave harsher sentences to defendants if they had been primed with black-associated words rather than neutral words, while judges who exhibited a black preference on the IAT gave less harsh sentences to defendants if they had been primed with black-associated words rather than neutral words. We did not find any significant relationship between the judges' IAT scores and either of the recidivism measures, although the data showed a similar trend.⁹⁹

To summarize, we found no overall difference between those judges primed with black-associated words and those primed with race-neutral words. This finding contrasts sharply with research conducted by Graham and Lowery, who found that police and parole officers primed with black-associated words were more likely than those primed with neutral words to make harsh judgments of juvenile offenders.¹⁰⁰ The officers who had seen the black-associated words

97 The results were as follows: Question 1, $z = 0.17$, $p = .87$; Question 2, $z = 0.09$, $p = .93$; and Question 3, $z = 1.62$, $p = .11$.

98 Our findings were: $z = 1.85$, $p = .06$.

99 For the first recidivism question, $z = 0.62$, $p = .53$; on the second recidivism question, $z = 0.54$, $p = .59$. As above, on these questions, the black judges and the white judges seemed to respond in similar ways. We ran the full model (predictors of prime, race of judge, IAT, and all interactions between these variables) on all three variables as well. Adding the race-of-judge terms and interactions did not produce any significant effects.

100 See Graham & Lowery, *supra* note 9, at 493–94, 496.

deemed the juveniles more culpable, more likely to recidivate, and more deserving of a harsh punishment.¹⁰¹

The overall lack of an effect of the racial prime, however, gives us little reason to conclude that the judges were not affected by their unconscious racial biases. We found in both the shoplifter case and the robbery case that judges who expressed a white preference on the IAT were somewhat more likely to impose harsher penalties when primed with black-associated words than when primed with neutral words, while judges who expressed a black preference on the IAT reacted in an opposite fashion to the priming conditions.

To be sure, we did not find a significant relationship between IAT scores and the judges' judgments of recidivism. That is, white preferences on the IAT did not lead judges primed with words associated with black Americans to predict higher recidivism rates. The judges made fairly race-neutral assessments of the two defendants' character. This result suggests that the correlation we found between IAT score and sentence might not be robust. But, of course, a judges' neutral assessment of character would be a small comfort to a juvenile defendant who received an excessive sentence due to his race.

2. Race Made Explicit

The fact that we did not explicitly provide any information about the race of the defendant (although judges obviously might have made assumptions about their race) is important because judges will commonly be aware of the race of the defendant appearing in front of them. To address this concern, we also gave our judges a hypothetical vignette in which we made race explicit. To enable comparison with another study, we used a vignette developed by Samuel Sommers and Phoebe Ellsworth.¹⁰²

We asked the judges to imagine they were presiding over a bench trial in which the prosecution charges André Barkley, a high school basketball player, with battering his teammate, Matthew Clinton. There is no question that Barkley injured Clinton, but Barkley claims, somewhat incredibly, that he was only acting in self-defense. We informed some of the judges that the defendant was an African American male and that the victim was a Caucasian male. We informed the

101 *Id.* Only police officers predicted that the defendant was more likely to recidivate; parole officers did not show any differences on this question. *Id.*

102 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, 216–17 (2001). We thank the authors for graciously sending us the materials and giving us permission to use them.

rest of the judges that the defendant was Caucasian and that the victim was African American. Following the scenario, we asked all of the judges to render a verdict and to rate their confidence in their judgment on a nine-point scale (from "Very Confident" to "Not at all Confident").¹⁰³

We found that the white judges were equally willing to convict the defendant whether he was identified as Caucasian or as African American. Among the white judges who read about an African American defendant, seventy-three percent (thirty-three out of forty-five) said they would convict, whereas eighty percent (thirty-five out of forty-four) of the white judges who read about a Caucasian defendant said that they would convict.¹⁰⁴ This contrasts sharply with the results obtained by Sommers and Ellsworth, who used only white participants. They found that ninety percent of the participants in their study who read about an African American defendant said that they would convict as compared to seventy percent of the participants who read about a Caucasian defendant.¹⁰⁵ On the other hand, we found that black judges were significantly more willing to convict the defendant when he was identified as Caucasian rather than as African American. When the defendant was identified as Caucasian, ninety-two percent (twenty-four out of twenty-six) of the black judges voted to convict; when he was identified as African American, however, only fifty percent (nine out of eighteen) voted to convict. The difference between the white judges and the black judges is statistically signifi-

103 We used the same question to elicit verdicts and confidence ratings as the one Sommers and Ellsworth used: "Based on the available evidence, if this were a bench trial, would you convict the defendant?" Below this were the words "Yes" and "No." Finally, we asked the judges, "How confident are you that your judgment is correct?" Below this question, the materials presented a nine-point scale, with "1" labeled "Not at all Confident" and "9" labeled "Very Confident." *Id.* at 217; *see also infra* Appendix A (providing the materials used in our study).

104 This difference was not statistically significant. Fishers exact test, $p = .62$.

105 The difference between our results and those obtained by Sommers and Ellsworth is significant: $\chi^2(1) = 6.74$, $p < .01$ (using the expected conviction rates of seventy percent for Caucasian defendants and ninety percent for African American defendants, as reported by Sommers & Ellsworth, Sommers & Ellsworth, *supra* note 102, at 217).

cant.¹⁰⁶ Analysis of the judges' assessments of their confidence in their verdicts produced similar results.¹⁰⁷

The focus of this study, however, is on the relationship between implicit bias and judgment. As above, we wanted to assess the effect of the interaction between the judges' IAT scores and the race of the defendant on the judges' verdicts. Unlike our results in the first study, however, we did not find even a marginally significant interaction here.¹⁰⁸ Judges who exhibited strong white preferences on the IAT did not judge the white and black defendants differently, and neither did judges who expressed black preferences on the IAT. Analysis of the confidence ratings produced the same result.¹⁰⁹

Because the white judges and the black judges reacted differently to the problem, we also conducted an analysis to account for these differences. To do this, we assessed the interaction between the race of the defendant and the IAT score, along with the race of the

106 The analysis consisted of a logistic regression of the verdict against the race of the defendant, the race of the judge, and the interaction of these two parameters. The interaction was significant, $z = 2.12$, $p = .03$, which was the result of the differential treatment of the two defendants by the black judges. The race of the defendant was also significant, $z = 2.81$, $p = .005$, indicating that overall, the judges were less likely to convict the African American defendant than the Caucasian defendant.

107 We combined the nine-point confidence measure with the binary outcome to create an eighteen-point scale. In our coding, a "1" corresponded to a judge who was very confident that the defendant should be acquitted, whereas an "18" corresponded to a judge who was very confident that the defendant should be convicted. The average confidence that the judges expressed in the defendant's guilt were as follows: white judges judging Caucasian defendants—13.64; white judges judging African American defendants—12.2; black judges judging Caucasian defendants—16.08; black judges judging African American defendants—9.89. Statistical analysis of these results (by ANOVA) produced results consistent with the analysis of the verdicts alone. That is, the judges were significantly more convinced of the Caucasian defendant's guilt than of the African American's guilt ($F(1, 129) = 15.04$, $p < .001$). This disparity was much more pronounced among black judges ($F(1, 129) = 5.84$, $p < .025$).

108 To accomplish this analysis, we conducted a logistic regression of the judges' verdict against the priming condition, the judges' IAT scores, and an interaction of the two. The interaction term reflects the effect of the IAT score on how the race of the defendant affected the judges' verdict. This term was not significant in the model, $z = 1.04$, $p = .30$.

109 We also replicated this analysis with the eighteen-point confidence ratings. See *infra* note 112. Specifically, we regressed the judges' confidence in the defendant's guilt against the defendant's race, the judges' IAT score, and the interaction between the race and IAT score. As with the verdict itself, this analysis showed that the race of the defendant was significant, $t\text{-ratio} = 3.49$, $p < .001$, but the interaction between race of defendant and IAT score was not, $t\text{-ratio} = 1.51$, $p = .13$.

judge.¹¹⁰ The three-way interaction between race of judge, race of defendant, and IAT score was significant.¹¹¹ This result means that the IAT scores of the black judges and the white judges had different effects on the judges' reactions to the race of the defendant, as we explain below in further analyses. Analysis of the confidence ratings produced similar results.¹¹²

To allow us to interpret this interaction, we ran the less complex analysis separately for black and white judges. That is, we assessed the interaction between the IAT score and race of the defendant in two separate analyses. With respect to the white judges, we found no significant results; if anything, the white judges with a greater white preference expressed a greater propensity to convict the Caucasian defendant rather than the African American defendant.¹¹³ Among black judges, however, those who expressed a stronger black preference on the IAT were less likely to convict the African American defendant relative to the Caucasian defendant.¹¹⁴ An analysis of confidence ratings produced similar results.¹¹⁵

The findings among black judges can best be seen by dividing the black judges into two groups: those who expressed a black preference on the IAT and those who expressed a white preference on the IAT. Among those black judges who expressed a black preference, one hundred percent (fourteen out of fourteen) voted to convict the Caucasian defendant, while only forty percent (four out of ten) of these

110 In this analysis, the race of the defendant and the interaction between race of judge and race of the defendant were significant, just as they were in the simpler models. (Race of defendant, $z = 1.99$, $p = .05$; interaction between race of the judge and race of the defendant, $z = 2.35$, $p = .02$. The interaction of the defendant's race and IAT score was not significant, $z = 1.00$, $p = .23$.)

111 The result was as follows: $z = 2.18$, $p = .03$.

112 Regressing the eighteen-point confidence rating against the race of the judge, the race of the defendant, the judges' IAT scores, and all interactions between these variables revealed significant effects for race of the defendant, $t\text{-ratio} = 2.95$, $p = .005$; a significant interaction of race of the defendant with race of the judge, $t\text{-ratio} = 2.68$, $p = .01$; and the three-way interaction of race of judge, race of defendant, and IAT score, $t\text{-ratio} = 2.68$, $p = .02$. The interaction of race of defendant and IAT scores was still not significant in this model, $t\text{-ratio} = 1.27$, $p = .20$.

113 The results are as follows: $z = 1.15$, $p = .25$.

114 The results are as follows: $z = 1.87$, $p = .06$. Given the high conviction rate of the black judges for the Caucasian defendant, this trend actually meant that they were more likely to convict the African American defendants to the extent that they exhibited greater white preferences on the IAT.

115 The white judges displayed a greater propensity to convict the Caucasian defendant relative to the African American defendant as the IAT score increased, but the trend did not approach significance, $t\text{-ratio} = 1.00$, $p = .40$. The black judges showed the opposite trend, which was significant: $t\text{-ratio} = 2.25$, $p = .03$.

judges voted to convict the African American defendant. Among those black judges who expressed a white preference, eighty-three percent (ten out of twelve) voted to convict the Caucasian defendant, while sixty-three percent (five out of eight) voted to convict the African American defendant. In effect, the black judges who expressed white preferences made verdict choices similar to those of their white colleagues, while black judges who expressed a black preference treated the African American defendant more leniently.

In sum, then, IAT scores predicted nothing among the white judges. Among the black judges, however, a black preference on the IAT was associated with a willingness to acquit the black defendant.

C. *Interpretation of Results*

Our research supports three conclusions. First, judges, like the rest of us, carry implicit biases concerning race. Second, these implicit biases can affect judges' judgment, at least in contexts where judges are unaware of a need to monitor their decisions for racial bias. Third, and conversely, when judges are aware of a need to monitor their own responses for the influence of implicit racial biases, and are motivated to suppress that bias, they appear able to do so.

Our first conclusion was perhaps the most predictable, though it is still troubling. Given the large number of Americans who have taken the IAT, and given the frequency with which white Americans display at least a moderate automatic preference for white over black, it would have been surprising if white judges had failed to exhibit the same automatic preference. Similarly, the black judges carry a more diverse array of implicit biases, just like black adults generally: some exhibit a white preference just like the white judges; others exhibit no preference; and some exhibit a black preference. Overall, like adults, most of the judges—white and black—showed a moderate-to-large degree of implicit bias in one direction or the other. If ordinary adults carry a “bigot in the brain,” as one recent article put it,¹¹⁶ then our data suggest that an invidious homunculus might reside in the heads of most judges in the United States, with the potential to produce racially biased distortions in the administration of justice.

It is worth noting, however, that the research on so-called “chronic egalitarians” suggests that this result was not inevitable. Some whites with longstanding and intense *personal* commitments to eradicating bias in themselves—chronic egalitarians—do not exhibit the preference for whites over blacks on the IAT that most white

116 Siri Carpenter, *Buried Prejudice: The Bigot in Your Brain*, SCI. AM. MIND, May 2008, at 32, 32.

adults show.¹¹⁷ Despite their *professional* commitment to the equal application of the law, judges do not appear to have the same habits of mind as the chronic egalitarians. The proportion of white judges in our study who revealed automatic associations of white with good and black with bad was, if anything, slightly higher than the proportion found in the online surveys of white Americans. Thus, a professional commitment to equality, unlike a personal commitment to the same ideal, appears to have limited impact on automatic racial associations, at least among the judges in our study. Alternatively, the overrepresentation of black Americans among the criminal defendants who appear in front of judges might produce invidious associations that overwhelm their professional commitment. In either case, our findings are consistent with the implicit associations found among capital defense attorneys. White capital defense attorneys, another group which might be expected to have strong professional commitments to the norm of racial equality,¹¹⁸ exhibit the same automatic preference for whites as the general population.¹¹⁹

Taken together, then, the research on judges and capital defense attorneys raises serious concerns about the role that unconscious bias might play in the criminal justice system. Jurors are drawn from randomly selected adults, and a majority of white jurors will harbor implicit white preferences. If police, prosecutors, jurors, judges, and defense attorneys all harbor anti-black preferences, then the system would appear to have limited safeguards to protect black defendants from bias. Based on IAT scores alone, both black judges and black jurors seem to be less biased than either white judges or white jurors, because black Americans show less implicit bias than white Americans. But even considerable numbers of blacks express implicit biases. Perhaps the only entity in the system that might avoid the influence of the bigot in the brain is a diversely composed jury.

That said, the rest of our results call into question the importance of IAT scores *alone* as a metric to evaluate the potential bias of decisionmakers in the legal system. Our second and third conclusions show that implicit biases *can* translate into biased decisionmaking under certain circumstances, but that they do not do so consistently.

117 See Gordon B. Moskowitz & Amanda R. Salomon, *Preconsciously Controlling Stereotyping: Implicitly Activated Egalitarian Goals Prevent the Activation of Stereotypes*, 18 SOC. COGNITION 151, 155 (2000).

118 See Theodore Eisenberg & Sheri Lynn Johnson, *Implicit Racial Attitudes of Death Penalty Lawyers*, 53 DEPAUL L. REV. 1539, 1540 (2004) (“One would hope that those who represent capital defendants (or at least African-American capital defendants) would themselves be free of racialized thinking . . .”).

119 *Id.* at 1546–48.

Implicit associations influenced judges—both black judges and white judges—when we manipulated the race of the defendant by subliminal methods. Judges with strong white preferences on the IAT made somewhat harsher judgments of the juvenile defendants after being exposed to the black subliminal prime, and judges with strong black preferences on the IAT were somewhat more lenient after exposure to the black subliminal prime. In effect, the subliminal processes triggered unconscious bias, and in just the way that might be expected.

The story for the explicit manipulation of race is more complicated, however. The white judges, unlike the white adults in the Sommers and Ellsworth study,¹²⁰ treated African American and Caucasian defendants comparably. But the proper interpretation of this finding is unclear. We observed a trend among the white judges in that the higher their white preference, the more *favorably* they treated the African American defendant in the battery case. Thus, among the white judges, implicit bias did not translate into racial disparities when the race of the defendant was clearly identified in an experimental setting.

We believe that the data demonstrate that the white judges were attempting to compensate for unconscious racial biases in their decisionmaking. These judges were, we believe, highly motivated to avoid making biased judgments, at least in our study. Codes of judicial conduct demand that judges make unbiased decisions, at least in our study.¹²¹ Moreover, impartiality is a prominent element in almost every widely accepted definition of the judicial role.¹²² Judges take these norms seriously. When the materials identified the race of the defendant in a prominent way, the white judges probably engaged in cognitive correction to avoid the appearance of bias.

The white judges in our study behaved much like the subjects in other studies who were highly motivated to avoid bias in performing an assigned task.¹²³ What made our white judges different from the subjects studied by these other researchers is that most of the judges reported that they suspected racial bias was being studied, despite the

120 See Sommers & Ellsworth, *supra* note 102, at 217.

121 See MODEL CODE OF JUDICIAL CONDUCT, at Canon 2 (2008) (“A judge shall perform the duties of judicial office impartially, competently, and diligently.”).

122 See, e.g., AM. BAR ASS’N, BLACK LETTER GUIDELINES FOR THE EVALUATION OF JUDICIAL PERFORMANCE, at Guideline 5-2.3 (2005), *available at* http://www.abanet.org/jd/lawyersconf/pdf/jpec_final.pdf (prescribing “[a]bsence of favor or disfavor toward anyone, including but not limited to favor or disfavor based upon race, sex, religion, national origin, disability, age, sexual orientation, or socioeconomic status”).

123 See Glaser & Knowles, *supra* note 39, at 171.

fact that the only cue they received was the explicit mention of the defendant's race.¹²⁴ We think this report was truthful, given that the judges behaved the same way as other white subjects who attempted to avoid the influence of implicit bias.

The black judges responded somewhat differently to the overt labeling of the defendant's race. Like the white judges, the black judges in our study also reported being aware of the subject of the study, yet they showed a correlation between implicit associations and judgment when race was explicitly manipulated. Among these judges, a greater white preference produced a greater propensity to convict the African American defendant. In other words, the black judges clearly reacted differently when they were conscious that race was being manipulated—a difference that correlated with their score on the race IAT.

We do not conclude, however, that black judges are less concerned about avoiding biased decisionmaking than white judges. We have no doubt that the professional norms against bias concern the black judges just as deeply as their white counterparts—if not more so. And we are mindful that research on the effect of race on judges' decisions in actual cases demonstrates no clear effects.¹²⁵ We believe that both white and black judges were motivated to avoid showing racial bias.

Why then did the black judges produce different results? We can only speculate, but we suspect that both groups of judges were keen to avoid appearing to favor the white defendant (or conversely, wanted to avoid appearing to disfavor the black defendant). Black judges, however, might have been less concerned with appearing to favor the black defendant than the white judges. Those black judges who expressed a white preference, however, behaved more like their white counterparts in this regard, thereby producing a correlation between verdict and IAT score among black judges.

We also cannot ignore the possibility that the judges were reacting to the race of the victim, rather than (or in addition to) the race of the defendant. In all cases, we identified the victim as the opposite

124 During our presentation, one of us asked for a show of hands to indicate how many thought we were studying race. While not the most ideal way to make this inquiry, and while we did not keep a precise count, most of the judges raised their hands.

125 See, e.g., Kathryn Abrams, *Black Judges and Ascriptive Group Identification*, in *NORMS AND THE LAW* 208, 215 (John N. Drobak ed., 2006) (“The most noteworthy feature of these studies is that they find no consistent, and only a few salient, differences in decisionmaking that correlate with the race of the judge.”).

race as the defendant. Furthermore, black judges might have reacted differently to the fact that the case involved a cross-racial crime.

Given our results, we cannot definitively ascribe continuing racial disparities in the criminal justice system to unconscious bias. We nevertheless can draw some firm conclusions. First, implicit biases are widespread among judges. Second, these biases can influence their judgment. Finally, judges seem to be aware of the potential for bias in themselves and possess the cognitive skills necessary to avoid its influence. When they are motivated to avoid the appearance of bias, and face clear cues that risk a charge of bias, they can compensate for implicit bias.

Whether the judges engage their abilities to avoid bias on a continual basis in their own courtrooms, however, is unclear. Judges are subject to the same significant professional norms to avoid prejudice in their courtrooms that they carried with them into our study. And judges might well point to our study as evidence that they avoid bias in their own courtrooms, where the race of defendants is often reasonably clear, and they never face subliminal cues. But courtrooms can be busy places that do not afford judges the time necessary to engage the corrective cognitive mechanisms that they seem to possess. And even though many decisions are made on papers only, judges might unwittingly react to names or neighborhoods that are associated with certain races. Control of implicit bias requires active, conscious control.¹²⁶ Judges who, due to time pressure or other distractions, do not actively engage in an effort to control the “bigot in the brain” are apt to behave just as the judges in our study in which we subliminally primed with race-related words. Moreover, our data do not permit us to determine whether a desire to control bias or avoid the appearance of bias motivates judges in their courtrooms the way it seemed to in our study.

Furthermore, judges might be overconfident about their abilities to control their own biases. In recently collected data, we asked a group of judges attending an educational conference to rate their ability to “avoid racial prejudice in decisionmaking” relative to other judges who were attending the same conference. Ninety-seven percent (thirty-five out of thirty-six) of the judges placed themselves in the top half and fifty percent (eighteen out of thirty-six) placed themselves in the top quartile, even though by definition, only fifty percent can be above the median, and only twenty-five percent can be in the

126 See Carpenter, *supra* note 116, at 37–38.

top quartile.¹²⁷ We worry that this result means that judges are overconfident about their ability to avoid the influence of race and hence fail to engage in corrective processes on all occasions.

To be sure, this is only one study, and it has its limitations. The results might be the product of the particular judges who participated in our study, or the materials we used, or even the fact that hypothetical scenarios were used. Most importantly, we cannot determine whether the mental processes of judges on the bench more closely resemble those of judges subliminally primed with race or those for whom race was explicitly manipulated. Thus, it is not clear how implicit racial bias influences judicial decisionmaking in court, but our study suggests, at a minimum, that there is a sizeable risk of such influence, so we turn in the next Part to reforms the criminal justice system might consider implementing.

IV. MITIGATING IMPLICIT BIAS IN COURT

To minimize the risk that unconscious or implicit bias will lead to biased decisions in court, the criminal justice system could take several steps. These include exposing judges to stereotype-incongruent models, providing testing and training, auditing judicial decisions, and altering courtroom practices. Taking these steps would both facilitate the reduction of unconscious biases and encourage judges to use their abilities to compensate for those biases.

A. *Exposure to Stereotype-Incongruent Models*

Several scholars have suggested that society might try to reduce the presence of unconscious biases by exposing decisionmakers to

127 These data were collected by us at a conference of New York City administrative law judges in the summer of 2008. As one of the questions, we asked the following:

Relative to the other judges attending this conference, how would you rate yourself on the following:

Avoiding racial bias in making decisions

___ In the highest quartile (meaning that you are more skilled at this than 75% of the judges attending this conference)

___ In the second highest quartile (meaning that you are more skilled at this than 50% of the judges in this room, but less skilled than 25% of the judges attending this conference)

___ In the second lowest quartile (meaning that you are more skilled at this than 25% of the judges in this room, but less skilled than 50% of the judges attending this conference)

___ In the lowest quartile (meaning that you are less skilled at this than 75% of the judges attending this conference).

stereotype-incongruent models.¹²⁸ This suggestion, in fact, probably represents the dominant policy proposal among legal scholars who write about unconscious bias.¹²⁹ We certainly agree, for example, that posting a portrait of President Obama alongside the parade of mostly white male judges in many courtrooms would be an inexpensive, laudable intervention.

Our results, however, also raise questions about the effectiveness of this proposal. The white judges from the eastern jurisdiction in our study showed a strong set of implicit biases, even though the jurisdiction consists of roughly half white judges and half black judges. Indeed, the level of implicit bias in this group of judges was only slightly smaller than that of the western jurisdiction, which included only two black judges (along with thirty-six white, five Latino, and two Asian judges). Exposure to a group of esteemed black colleagues apparently is not enough to counteract the societal influences that lead to implicit biases.

Consciously attempting to change implicit associations might be too difficult for judges. Most judges have little control over their dockets, which tend to include an overrepresentation of black criminal defendants.¹³⁰ Frequent exposure to black criminal defendants is apt to perpetuate negative associations with black Americans. This exposure perhaps explains why capital defense attorneys harbor negative associations with blacks,¹³¹ and might explain why we found slightly greater negative associations among the white judges than the population as a whole (although as we noted above, the latter finding might have other causes).

B. Testing and Training

The criminal justice system might test candidates for judicial office using the IAT or other devices to determine whether they possess implicit biases. We do not suggest that people who display strong

128 Jolls & Sunstein, *supra* note 4, at 988–90; Kang & Banaji, *supra* note 7, at 1105–08.

129 See, e.g., Kang & Banaji, *supra* note 7, at 1112 (“In *Grutter v. Bollinger*, the Court emphasized that student diversity was valuable because it could help ‘break down racial stereotypes.’” (quoting *Grutter v. Bollinger*, 539 U.S. 306, 330 (2003))); see also Kang, *supra* note 8, at 1579–83 (arguing that public broadcasting should be regulated so as to promote positive images of minorities).

130 BUREAU OF JUSTICE STATISTICS, U.S. DEP’T OF JUSTICE, FELONY DEFENDANTS IN LARGE URBAN COUNTIES, 2004, at 1 (2004), available at <http://www.ojp.usdoj.gov/bjs/pub/pdf/fdluc04.pdf> (stating that an estimated forty percent of defendants were black).

131 See Eisenberg & Johnson, *supra* note 118, at 1553–56.

white preferences on the IAT should be barred from serving as judges, nor do we even support using the IAT as a measure of qualification to serve on the bench.¹³² The direct link between IAT score and decisionmaking is far too tenuous for such a radical recommendation. And our data show that judges can overcome these implicit biases at least to some extent and under some circumstances. Rather, knowing a judge's IAT score might serve two other purposes. First, it might help newly elected or appointed judges understand the extent to which they have implicit biases and alert them to the need to correct for those biases on the job.¹³³ Second, it might enable the system to provide targeted training about bias to new judges.¹³⁴

Judicial training should not end with new judges, however. Training for sitting judges is also important. Judicial education is common these days, but one problem with it, at least as it exists at this time, is that it is seldom accompanied by any testing of the individual judge's susceptibility to implicit bias, or any analysis of the judge's own decisions, so the judges are less likely to appreciate and internalize the risks of implicit bias.¹³⁵ As Timothy Wilson and his colleagues have observed, "people's default response is to assume that their judgments are uncontaminated."¹³⁶ Surely this is true of judges as well. Moreover, because people are prone to egocentric bias, they readily assume that they are better than average, or the factors that might induce others to make poor or biased decisions would not affect their own decisions. Our research demonstrates that judges are inclined to make the same sorts of favorable assumptions about their own abilities that non-judges do.¹³⁷ Therefore, while education regarding implicit bias as a general matter might be useful, specific training revealing the vulnerabilities of the judges being trained would be more useful.¹³⁸

Another problem with training is that although insight into the direction of a bias frequently can be gained, insight into the magni-

132 Others have made tentative suggestions that the IAT be used as a screening device for certain professions. See, e.g., IAN AYRES, *PERVASIVE PREJUDICE?* 424 (2001) ("Implicit attitude testing might also itself be used as a criterion for hiring both governmental and nongovernmental actors.").

133 Green et al., *supra* note 10, at 1237 ("These findings support the IAT's value as an educational tool.").

134 See *id.* (recommending "securely and privately administered IATs to increase physicians' awareness of unconscious bias").

135 See Carpenter, *supra* note 116, at 32.

136 Timothy D. Wilson et al., *Mental Contamination and the Debiasing Problem*, in *HEURISTICS AND BIASES* 185, 190 (Thomas Gilovich et al. eds., 2002).

137 See Guthrie et al., *Judicial Mind*, *supra* note 58, at 814–15.

138 See Green et al., *supra* note 10, at 1237.

tude of that bias cannot. One group of psychologists provided the following example:

Consider Ms. Green, a partner in a prestigious law firm, who is interviewing candidates for the position of an associate in her firm. When she interviews Mr. Jones, a young African-American attorney, she has an immediate negative impression, finding him to be arrogant and lacking the kind of brilliance she looks for in new associates. Ms. Green decides that her impression of Mr. Jones was accurate and at a meeting of the partners, argues against hiring him. She wonders, however, whether her negative evaluation was influenced by Mr. Jones' race.¹³⁹

The psychologists explained:

Ms. Green may know that her impression of Mr. Jones is unfairly negative and want to avoid this bias, but have no idea of the extent of the bias. Should she change her evaluation from "Should not be hired" to "Barely acceptable" or to "Best applicant I've seen in years"?¹⁴⁰

This scenario illustrates the problem well. How is one to know if correction is warranted, and if so, how much?¹⁴¹ In a circumstance like the one depicted above or like any of the circumstances described in the materials included in our study, there is a risk of insufficient correction, unnecessary correction, or even overcorrection, resulting in a decision that is distorted as a result of the adjustment, but simply in the opposite direction.¹⁴² Testing might mitigate this problem by

139 Wilson et al., *supra* note 136, at 185.

140 *Id.* at 187.

141 *See id.* at 191 ("Three kinds of errors have been found: *insufficient correction* (debiasing in the direction of accuracy that does not go far enough), *unnecessary correction* (debiasing when there was no bias to start with), and *overcorrection* (too much debiasing, such that judgments end up biased in the opposite direction).").

142 *See id.* (suggesting that people's "corrected judgments might be worse than their uncorrected ones"); *see also* Antony Page, *Batson's Blind-Spot: Unconscious Stereotyping and the Peremptory Challenge*, 85 B.U. L. REV. 155, 239-40 (2005) ("One major problem for any correction strategy is determining the magnitude of the correction required. Unfortunately, people are not very good at this determination. Some research suggests that among those who are very motivated to avoid discrimination, overcorrection is a common problem. . . . A second problem is that a correction strategy appears to require significant cognitive resources . . ." (citations omitted)); *id.* at 241-42 ("[T]o consciously and willfully regulate one's own . . . evaluations [and] decisions . . . requires considerable effort and is relatively slow. Moreover, it appears to require a limited resource that is quickly used up, so conscious self-regulatory acts can only occur sparingly and for a short time.") (omissions in original) (quoting John A. Bargh & Tanya L. Chartrand, *The Unbearable Automaticity of Being*, 54 AM. PSYCHOL. 462, 476 (1999)).

helping judges appreciate how much compensation or correction is needed.

The results of our study are thus somewhat surprising in that the white judges' corrections in the case in which the defendant's race was explicit seemed to be neither too much nor too little. On average, these judges treated white and black defendants about the same. This result cannot, however, reasonably be taken as meaning that judges correct for the influence of implicit bias perfectly in all cases in which they attempt to do so. We presented only one scenario—other cases might produce overcompensation or undercompensation. And individual judges are apt to vary in terms of their willingness or ability to correct for the influence of unconscious racial bias. Also, the white judges were slightly less harsh on the black defendants. The difference simply failed to rise to the level of statistical significance, as it was small (only six percentage points). Had we collected data on a thousand judges rather than a hundred, we might have begun to observe some overcompensation or undercompensation.

C. Auditing

The criminal justice system could also implement an auditing program to evaluate the decisions of individual judges in order to determine whether they appear to be influenced by implicit bias. For example, judges' discretionary determinations, such as bail-setting, sentencing, or child-custody allocation, could be audited periodically to determine whether they exhibit patterns indicative of implicit bias. Such proposals have been suggested as correctives for umpires in Major League Baseball and referees in the National Basketball Association after both groups displayed evidence of racial bias in their judgments.¹⁴³

Auditing could provide a couple of benefits. First, it would obviously increase the available data regarding the extent to which bias affects judicial decisionmaking. Second, it could enhance the accountability of judicial decisionmaking.¹⁴⁴ Unfortunately, judges operate in an institutional context that provides little accountability, at least in the sense that they receive little prompt and useful feed-

143 See Christopher A. Parsons et al., *Strike Three: Umpires' Demand for Discrimination* 24–25 (Nat'l Bureau of Econ. Research, Working Paper Series, Paper No. 13665, 2007), available at <http://ssrn.com/abstract=1077091>; Joseph Price & Justin Wolfers, *Racial Discrimination Among NBA Referees* 30 (Nat'l Bureau of Econ. Research, Working Paper Series, Paper No. 13206, 2007), available at <http://ssrn.com/abstract=997562>.

144 Accountability improves performance in other contexts, so it likely would do so for judges as well. See Jennifer S. Lerner & Philip E. Tetlock, *Accounting for the Effects of Accountability*, 125 *PSYCHOL. BULL.* 255, 270–71 (1999).

back.¹⁴⁵ Existing forms of accountability, such as appellate review or retention elections, primarily focus on a judge's performance in a particular case, not on the systematic study of long-term patterns within a judge's performance that might reveal implicit bias.¹⁴⁶

D. *Altering Courtroom Practices*

In addition to providing training or implementing auditing programs, the criminal justice system could also alter practices in the courtroom to minimize the untoward impact of unconscious bias. For example, the system could expand the use of three-judge courts.¹⁴⁷ Research reveals that improving the diversity of appellate court panels can affect outcomes. One study found that "adding a female judge to the panel more than doubled the probability that a male judge ruled for the plaintiff in sexual harassment cases . . . and nearly tripled this probability in sex discrimination cases."¹⁴⁸ In trial courts, judges typically decide such issues alone, so adopting this mechanism would require major structural changes. Although convening a three-judge trial court was once required by statute when the constitutionality of a state's statute was at issue,¹⁴⁹ three-judge trial courts are virtually nonexistent today.¹⁵⁰ The inefficiency of having three judges decide cases that one judge might be able to decide nearly as well led to their demise, and this measure might simply be too costly to resurrect.

Another possibility would be to increase the depth of appellate scrutiny, such as by employing *de novo* review rather than clear error review, in cases in which particular trial court findings of fact might be tainted by implicit bias. For example, there is some evidence that male judges may be less hospitable to sex discrimination claims than they ought to be.¹⁵¹ If that bias does exist, less deferential appellate review by a diverse panel might offer a partial solution.

145 See Guthrie et al., *How Judges Decide*, *supra* note 58, at 32.

146 See, e.g., Jean E. Dubofsky, *Judicial Performance Review: A Balance Between Judicial Independence and Public Accountability*, 34 *FORDHAM URB. L.J.* 315, 320–22 (2007) (explaining that the judicial performance review system in Colorado focuses only on a judge's performance in a particular case).

147 See Michel E. Solimine, *Congress, Ex Parte Young, and the Fate of the Three-Judge District Court*, 70 *U. PITT. L. REV.* 101, 128–134 (2008).

148 Jennifer L. Peresie, Note, *Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts*, 114 *YALE L.J.* 1759, 1778 (2005).

149 Note, *Judicial Limitation of Three-Judge Court Jurisdiction*, 85 *YALE L.J.* 564, 564 (1976).

150 Arthur D. Hellman, *Legal Problems of Dividing a State Between Federal Judicial Circuits*, 122 *U. PA. L. REV.* 1188, 1225 (1974).

151 See Peresie, *supra* note 148, at 1778.

CONCLUSION

Our study contains both bad news and good news about implicit biases among judges. As expected, we found that judges, like the rest of us, possess implicit biases. We also found that these biases have the potential to influence judgments in criminal cases, at least in those circumstances where judges are not guarding against them. On the other hand, we found that the judges managed, for the most part, to avoid the influence of unconscious biases when they were told of the defendant's race.

The presence of implicit racial bias among judges—even if its impact on actual cases is uncertain—should sound a cautionary note for those involved in the criminal justice system. To prevent implicit biases from influencing actual cases, we have identified several reforms that the criminal justice system could implement, ranging from relatively inexpensive measures, like implementing focused judicial training and testing, to relatively expensive measures, like altering courtroom practices. To render justice blind, as it is supposed to be, these reforms are worth considering.

APPENDIX A: MATERIALS

Shoplifter Case

You are presiding over a case involving criminal charges against a juvenile, William T. William is a 13-year-old who was arrested for shoplifting in a large, upscale toy store in _____. He has no prior record. You are trying to get a sense of the case and the only facts available to you follow:

According to a store clerk, on Saturday, April 2, at about two o'clock in the afternoon, the clerk observed William putting videogames under his shirt. The clerk rang for a security guard, but before the guard arrived, the boy started to leave the store. When the clerk grabbed William, the boy dropped the toys and kicked him in an attempt to escape. A uniformed security guard arrived as the clerk let go of William, and when the guard told the boy to stop, he did.

According to the security guard, when he arrived he observed five items on the floor in front of William. The prices of those items together added up to \$90. He said that William told him that he was shopping, and showed him \$10 he had brought along with which to make purchases. William claimed that he had used his shirt as a sort of pouch to hold the items he was looking at. William also told the guard he was startled when grabbed by someone from behind, and then tripped, but that he did not kick anyone.

1. In your opinion, without regard to the options actually available in this kind of situation, what would be the most appropriate disposition of this case?

- ___ 1) Dismiss it with an oral warning
- ___ 2) Adjourn the case in contemplation of dismissal
(assuming William gets in no further trouble)
- ___ 3) Put William on probation for six months or less
- ___ 4) Put William on probation for more than six months
- ___ 5) Commit William to a juvenile detention facility for six months or less
- ___ 6) Commit William to a juvenile detention facility for more than six months
- ___ 7) Transfer William to adult court

2. In your opinion, on a scale of one to seven, how likely is it that William will later commit a crime similar to the one with which he is charged?

Very Likely Not at all Likely
 1 2 3 4 5 6 7

3. In your opinion, on a scale of one to seven, how likely is it that William will commit more serious crimes in the future?

Very Likely

1

2

3

4

5

6

7

Not at all Likely

Robbery Case

You are presiding over a case involving criminal charges against a juvenile, Michael S., who was arrested for armed robbery of a gas station when he was two days shy of his seventeenth birthday. He has one prior arrest for a fight in the school lunchroom the previous year. You are trying to get a sense of the case and the only facts available to you follow:

According to the gas station clerk, on Friday, March 17, at about seven in the evening, she heard a male voice say, "Don't look at me, but give me the money." She kept her eyes down, and as she opened the cash register, the man said, "I could shoot you, don't think I won't." She handed him the drawer's contents (\$267.60) and saw him run out the door with a gun. After he jumped into the passenger side of a car and it left, she called the police.

According to the responding officer, the clerk could not identify the robber, but a customer said he thought he recognized Michael, and gave the officer Michael's name and address. Michael's mother was home, and at nine forty-five, Michael walked in the door, was given Miranda warnings, and waived his rights. He first stated that he had just been hanging around with friends, not doing anything special. After the officer asked who the friends were, Michael admitted that he had walked into the gas station with a gun. He told the officer that he said to the clerk, "Give me the money, please. I don't want to hurt you." Michael insisted that the gun was not loaded and that he no longer had it. He said that the money was gone, that he was sorry, and would pay it back. When asked why he did it, Michael said that his friends had dared him, but he would not reveal who those friends were, or to whom the gun belonged.

1. In your opinion, without regard to the options actually available in this kind of situation, what would be the most appropriate disposition of this case?

- _____ 1) Dismiss it with an oral warning
- _____ 2) Adjourn the case in contemplation of dismissal (assuming Michael gets in no further trouble)
- _____ 3) Put Michael on probation for six months or less
- _____ 4) Put Michael on probation for more than six months

- _____ 5) Commit Michael to a juvenile detention facility for six months or less
- _____ 6) Commit Michael to a juvenile detention facility for more than six months
- _____ 7) Transfer Michael to adult court

2. In your opinion, on a scale of one to seven, how likely is it that Michael will later commit a crime similar to the one with which he is charged?

Very Likely Not at all Likely
 1 2 3 4 5 6 7

3. In your opinion, on a scale of one to seven, how likely is it that Michael will commit more serious crimes in the future?

Very Likely Not at all Likely
 1 2 3 4 5 6 7

Battery Case

Defendant: André Barkley, 6’0”, 175 lbs., African American male, 18 years old, student

Alleged Victim: Matthew Clinton, 6’2”, 185 lbs., Caucasian male, 16 years old, student

Charge: One Count of Battery with Serious Bodily Injury

Prosecution

The prosecution claims that André Barkley is guilty of battery with serious bodily injury. Barkley was the starting point guard on the high school basketball team, but the team had been struggling, and the coach decided to bench him in favor of a younger, less experienced player named Matthew Clinton. Before the first game after the lineup change, Barkley approached Clinton in the locker room and began yelling at him. Witnesses explain that the frustrated defendant told Clinton, “You aren’t half the player I am, you must be kissing Coach’s ass pretty hard to be starting.”

When other teammates stepped between the two players, Barkley told them to get out of the way. When two other players then grabbed Barkley and tried to restrain him, the defendant threw them off, pushed Clinton into a row of lockers, and ran out of the room, according to prosecution witnesses. As a result of this fall, two of Clinton’s teeth were chipped and he was knocked unconscious. The prosecution claims that Barkley has shown no remorse for his crime, and has even expressed to friends that Clinton “only got what he had coming.”

Defense

The defense claims that Barkley was merely acting in self-defense, and that Clinton's injuries were accidental. According to an assistant coach, Barkley did not get along with many people on the team and had been the subject of obscene remarks and unfair criticism from many of his teammates throughout the season. Barkley claims that he was afraid for his own safety during the altercation in the locker room and "definitely felt ganged up on."

Barkley admits he "might have been aggressive towards Matthew and started the whole thing," but says that he was just frustrated and the argument was "nothing that should have started a big locker room fight or anything." Barkley claims that when several other players grabbed him from behind for no reason, he tried to break free and must have accidentally knocked into Clinton in the attempt to get out of the locker room. He explained that the reason he never apologized to Clinton in the hospital was that he "didn't think he'd want to see me," but Barkley did say he "was truly, truly sorry" that Clinton had been injured.

1. Based on the available evidence, if this were a bench trial, would you convict the defendant?

Yes

No

2. How confident are you that your judgment is correct?

Very Confident

Not at all Confident

1

2

3

4

5

6

7

8

9

Demographic Questions Provided to Judges

What is the title of the judicial position you currently hold?

How many years have you served as a Judge (in any position)?

___ years

Please identify your gender:

___ male

___ female

During your judicial career, approximately what percentage of your time has been devoted to the following areas:

___ Criminal cases

___ Civil cases

___ Family law cases

___ Probate or trusts

___ Other

Which of the two major political parties in the United States most closely matches your own political beliefs?

- The Republican Party
- The Democratic Party

Please identify your race (Check all that apply)

- White (non-Hispanic)
- Black or African American
- Hispanic or Latino
- Asian
- Native American or Pacific Islander
- Other

APPENDIX B: IAT PROCEDURE

We used seven rounds of trials to produce the IAT score. Rounds one, two, three, five, and six are essentially practice rounds designed to minimize order effects and variation associated with unfamiliarity with the task. The study begins with one round in which the participants only sort black and white faces. In this round the word “White” appeared in the upper left and the word “Black” appeared in the upper right of the screen. In each trial, one of ten faces, five white and five black, appeared in the middle of the screen.¹⁵² The faces appeared at random, although an equal number of white and black faces appeared in the sixteen trials.¹⁵³

The instructions before each round informed the judges as to what they would be sorting in the upcoming round. For example, in the first round, the instructions indicated that the judge should press the “E” key (labeled with a red dot) if a white face appeared and the “I” key (also labeled with a red dot) if a black face appeared. The materials also state that if the judge pressed the correct key, the next face would appear; if the judge pressed the wrong key, a red “X” would appear. These instructions were similar in all seven rounds of the IAT.¹⁵⁴

The remaining six rounds were similar to the first, although they varied the stimuli and categories. In the second round, instead of the

152 The faces were taken from the Project Implicit website. See Brian A. Nosek et al., Project Implicit, Stimulus Materials (2006), <http://www.projectimplicit.net/stimuli.php>. They include only the center of the face, with ears, hair, and anything below the chin cropped out. None of the faces has facial hair, eyeglasses, or distinguishing features. *Id.* (providing faces that can be downloaded under the “race faces” stimulus set).

153 In this respect we varied from the procedures recommended by Greenwald and his colleagues, see Greenwald et al., *supra* note 29, at 198, by reducing the practice rounds from the twenty they suggested to sixteen. We did this in the interest of saving time. We did retain the forty trials in the critical rounds. We had more time available in the western jurisdiction, and increased the length of rounds three and six to twenty trials.

154 The exact instructions were as follows:

In the first round, the two CATEGORIES that you are to distinguish are:
BLACK vs. WHITE faces.

Press the “E” key if the TARGET is a WHITE face.

Press the “I” key if the TARGET is a BLACK face.

Remember that an “X” will appear when you make an error. Whenever the “X” appears, correct the mistake by pressing the other key.

Please respond AS RAPIDLY AS POSSIBLE, but don’t respond so fast that you make many errors. (Occasional errors are okay.)

Press the space bar when you are ready to begin.

black and white faces, the computer presented good and bad words. These consisted of seven words with positive associations (Joy, Love, Peace, Wonderful, Pleasure, Friend, Laughter, Happy) and seven words with negative associations (Agony, Terrible, Horrible, Nasty, Evil, War, Awful, Failure). Like the faces, these words were taken from previous work on the IAT. Throughout the trials in the second round, the word "Good" remained in the upper-left of the computer screen and the word "Bad" remained in the upper-right of the computer screen. The judges were instructed in a similar fashion to round one, to press the "E" key when a good word appeared in the center of the screen and to press the "I" key when a bad word appeared in the center of the screen.

The third round combined the tasks in the first two rounds. The words "White or Good" appeared in the upper-left of the computer screen and the words "Black or Bad" appeared in the upper-right of the computer screen. Thus, the task presented both categories in the same spatial location as they had been in the first two rounds. The instructions indicated to the judge that either a white or black face or a good or bad word would appear in the center of the computer screen. The instructions continued that the judges should press the "E" key if either a white face or a good word appeared and the "I" key if either a black face or a bad word appeared. Although the computer selected randomly from the faces and concept words, the computer presented an equal number of names and faces of both types. We presented the judges with sixteen trials of this task

Round four was identical to round three in every respect except that the computer presented forty trials, rather than sixteen.

Round five prepared the judges for the reverse association. To create the reversal, the spatial locations of the good and bad words were reversed. The word "Bad" was moved to the left and the word "Good" was moved to the right. The fifth round was thus identical to the second round in that the computer presented only the good and bad words, but that the computer presented the words in their new locations. The instructions were also identical to those of round two except that they identified the new locations and corresponding response keys for the words.

The penultimate round paired the good and bad words in their new locations with the black and white labels in their original location. Thus, the words "White or Bad" appeared in the upper left and the words "Black or Good" appeared in the upper right. The instructions resembled those for rounds three and four. They indicated, however, that judges should press the "E" key if a white face or bad word appeared and to press the "I" key if a black face or good word

appeared. Round six, like the other practice rounds, consisted of sixteen trials.

Round seven was identical to round six in every respect except that the computer presented forty trials, rather than sixteen. The computer recorded the reaction times between the presentation of the stimuli and the time of the correct response for all judges in all rounds. The computer also recorded which stimuli it presented and whether an error occurred.

APPENDIX C: IAT SCORING

Scoring the IAT requires researchers to make several judgments about the data. It requires deciding which of the seven rounds to use (some studies make use of the practice rounds); how to manage latencies that seem too long or too short; how to assess erroneous responses; how to identify and score participants who respond too slowly, too quickly, or made too many errors; whether to standardize the responses; and whether to use every round in a trial (or drop the first two, which commonly produce excessively long latencies). Greenwald and his colleagues tested essentially all variations on answers to these issues and produced a scoring method that they believe maximizes the correlation between the IAT and observed behavior.¹⁵⁵

We used two different scoring methods. First, for each judge, we calculated the difference between the average latency in the stereotype-congruent rounds in which the judges sorted white/good versus black/bad and the average latency in the stereotype-incongruent rounds in which the judges sorted white/bad versus black/good. This procedure follows the method that other researchers have used in reporting data from hundreds of thousands of participants collected on the Internet.¹⁵⁶ Hence, we can compare this average score with that of large groups of ordinary adults. (We describe this procedure at greater length below.)

In an exhaustive review of IAT methodology, however, Greenwald and his colleagues concluded that the average difference might not be the best measure of implicit associations.¹⁵⁷ These researchers found that people who are slower on the task produce larger differences in their IAT scores.¹⁵⁸ This tendency confounds the IAT score, as people who are simply less facile with a keyboard will appear to have stronger stereotypic associations. Furthermore, Greenwald and his colleagues also found that the average difference did not correlate as well with people's decisions and behavior as other scoring methods.¹⁵⁹ After conducting their review, Greenwald and his colleagues identified a preferred scoring method, which we followed to assess the correlation between IAT effects and judges' decisions.¹⁶⁰ The method essentially uses the mean difference for each participant divided by the standard

155 Greenwald et al., *supra* note 29, at 212–15.

156 Nosek et al., *supra* note 17, at 103–04.

157 Greenwald et al., *supra* note 29, at 212–15.

158 *Id.* at 201–02.

159 *Id.* at 203.

160 *Id.* at 214 tbl.4.

deviation of that participant's response latencies, although it includes some variations. (We also describe this procedure at greater length below.)

1. *Mean-Difference IAT Score Calculation*

To calculate the mean-difference IAT score, we largely followed the procedures outlined in Nosek and his colleagues' report of IAT scores from tens of thousands of people collected through the Internet.¹⁶¹ We also wanted to compare our results with the more detailed, contemporary Internet data collected and reported on the "Project Implicit" website, which appears to use the same scoring method.¹⁶² Because the data in these studies come from voluntary participants who access the site on the Internet, the authors have adopted a number of techniques for excluding data from participants who may have wandered off during the study or are otherwise not fully engaged with the tasks.¹⁶³ While such techniques are less appropriate for our participants, who were engaged in person, we followed the Project Implicit scoring methods to facilitate a comparison.

The authors of the Internet study first adjusted raw latency scores that seemed much slower or faster than participants who are fully engaged with the task. The researchers treat any latency larger than 3000 milliseconds (ms) as 3000 ms, and any latency shorter than 300 ms as 300 ms.¹⁶⁴ The researchers also eliminated the first two trials in all rounds from consideration, having found that these rounds often displayed an erratic pattern of long latencies—presumably because participants commonly begin the task, and then pause to get settled in.¹⁶⁵ These researchers also excluded participants who failed to perform to certain criteria. They excluded participants who exhibited overall average latencies in the two critical rounds greater than 1800 ms, or who displayed average latencies in either of the two critical rounds (four or seven) greater than 1500 ms.¹⁶⁶ They also excluded participants who produced any critical round in which more than twenty-five percent of the latencies were less than 300 ms.¹⁶⁷ Finally, they excluded participants who made more than ten errors in any crit-

161 Nosek et al., *supra* note 17, at 103–04.

162 Project Implicit, Background Information (2002), <https://implicit.harvard.edu/implicit/demo/background/index.jsp> (last visited on Mar. 9, 2009).

163 See Nosek et al., *supra* note 17, at 104.

164 *Id.*

165 *Id.*

166 *Id.*

167 *Id.*

ical round.¹⁶⁸ These researchers report that these criteria resulted in the exclusion of fifteen percent of their subjects.¹⁶⁹ After these adjustments and exclusions, these researchers calculated the mean difference between the critical stereotype-congruent round (either round four or seven) and the stereotype-incongruent rounds (either round four or seven).¹⁷⁰

We followed these procedures to calculate the mean IAT score for the judges in our study. We capped latencies greater than 3000 ms as 3000 ms, and raised latencies lower than 300 ms to 300 ms.¹⁷¹ We also discarded the first two rounds from the analysis. We excluded the results of the race IAT from six judges (or 4.5%) who produced either mean latencies greater than 1800 ms in one of the two critical rounds of the race IAT or a mean across both rounds greater than 1500 ms.¹⁷² Similarly, we excluded the results of the gender IAT from ten judges (or 7.5%) who violated one or both of these criteria.¹⁷³ Nosek and his colleagues reported that they eliminated two percent of their participants for being too slow,¹⁷⁴ whereas we eliminated more. At the same

168 *Id.*

169 *Id.*

170 *Id.*

171 None of the judges provided latencies that were less than 300 ms in either of the two critical rounds measuring the race IAT; two of the judges provided responses that were faster than 300 ms in the gender IAT (one round each). Many more of the judges produced latencies that exceeded 3000 ms. On the race IAT, fifty-eight judges (or 50.4%) produced at least one latency greater than 3000 ms in the stereotype-congruent round (round four). Specifically, in the stereotype-congruent round: thirty-three judges produced one long latency; twenty produced two; three produced three; and two produced four. In the stereotype-incongruent round on the race IAT (round seven), sixty-eight judges (or 59.1%) produced at least one latency greater than 3000 ms. Specifically, in the stereotype-incongruent round: thirty-three judges produced one long latency; twelve produced two; ten produced three; four produced four; two produced five; four produced six; and three produced seven. On the gender IAT, fifty-seven judges (or 49.6%) produced at least one latency greater than 3000 ms in the stereotype-congruent round (round seven). Specifically, in the stereotype-congruent round: thirty-six judges produced one long latency; seven produced two; nine produced three; three produced four; one produced five; and one produced eight. In the stereotype-incongruent round on the gender IAT (round four), fifty-six judges (or 48.7%) produced at least one latency greater than 3000 ms. Specifically, in the stereotype-incongruent round: twenty-seven judges produced one long latency; fifteen produced two; six produced three; three produced four; two produced five; one produced six; and one produced seven. Note that because some of these long latencies fell into the first two rounds, they are not included in the analysis.

172 One of the judges violated both criteria. We calculated both means after excluding the first two rounds.

173 Four judges violated both criteria.

174 Nosek et al., *supra* note 17, at 104.

time, none of the judges in our studies produced more than a twenty-five percent error rate in either of the critical rounds in either IAT. By contrast, Nosek and his colleagues eliminated roughly thirteen percent of their participants for having high error rates.¹⁷⁵ The judges were thus slower and more accurate than Nosek and his colleagues' subjects, and overall, the application of their criteria eliminated fewer judges than their results would have predicted.

Unlike Nosek and his colleagues,¹⁷⁶ we did not randomize the order in which we presented the IAT. That is, roughly half of the participants in the Internet sample receive the stereotype-congruent round first, while half receive the stereotype-incongruent round first. The seven-round IAT is designed to reduce order effects substantially, but nevertheless, they remain. Greenwald and his colleagues report that the IAT scores can correlate weakly with the order in which the materials are presented.¹⁷⁷ Randomizing the order would have produced a cleaner measure of the IAT effect across all judges, but would have reduced the correlation between the IAT score and behavior.¹⁷⁸ Hence, all of our judges received the materials in the same order. On the race IAT, judges receive the stereotype-congruent pairing first (white/good and black/bad) and on the gender IAT, judges receive the stereotype-incongruent pairing first (male/humanities and female/science). Our procedure would have tended to increase the IAT score on the race IAT, as compared to the sample by Nosek and his colleagues, and decrease the IAT score on the gender IAT.

By using these procedures, we scored judges in exactly the same method as Nosek and his colleagues in the data that they harvested

175 *Id.*

176 *Id.*

177 Greenwald et al., *supra* note 29, at 210 tbl.2, report the effect of order with a correlation coefficient, rather than a mean or percent difference. They report that the correlation varies with the IAT, noting that the gender IAT that we used here produces a higher correlation between order and IAT score than do other IATs. They report correlations as high as 0.29 (depending upon the scoring method), which would mean that order can account for up to ten percent of the IAT score. *Id.* By contrast, the race IAT that we used produces small correlations with order, ranging from 0.002 to 0.054; thus, order accounts for, at most, one-quarter of one percent of the IAT score. The order effects seem to vary with context, and hence we cannot be certain of the extent of the influence of order on our materials.

178 Had we randomized the order, each judge's IAT score would have varied with the order to some extent. This would have introduced some variation to the IAT score that would inherently reduce the correlation we observed across all judges. Our measure of the IAT score across all judges would have been more reliable had we randomized, but the IAT score for the individual judges would have been less consistent, thereby interfering with the correlation.

from the Internet. Because laboratory data are obviously different in some respects, we only treated the data this way for purposes of comparison with the Internet samples, and not for assessing the correlation between the IAT scores and the decisions that judges made. For the correlations, we calculated a standardized score.

2. *Standardized IAT Score Calculation*

To calculate the standardized IAT score, we followed the procedures recommended by Greenwald and his colleagues.¹⁷⁹ These researchers designed their methods precisely to improve the reliability and predictive power of their measures.¹⁸⁰ We use the methods that produced the highest correlations between implicit measures and behavioral measures. They differ from the scoring method used to calculate the mean differences. As noted above, we used the Greenwald methodology to collect the IAT scores.¹⁸¹ Following those scoring procedures, we removed single trials with latencies greater than 10,000 ms (that is, ten seconds) from the analysis. We otherwise left low and high values in the analysis without adjustment. We made no correction for errors, because our IAT collection methods required the judges to provide the correct response before proceeding and hence the latency includes the delay that would result from an incorrect answer. Error rates were also low, as noted above. Following Greenwald and his colleagues' scoring method, we used all of the trials, rather than dropping the first two in the round.

We departed from the method Greenwald and his colleagues endorse, however, in one respect. Those researchers suggested using the two paired practice rounds (rounds three and six) in the analysis.¹⁸² They reported that using this data produced slightly higher correlations between the IAT scores and explicit choices.¹⁸³ We found, however, that latencies in the practice rounds were highly erratic. A high percentage of the trials eliminated for being greater than 10,000 ms were in the trial rounds.¹⁸⁴ Even with these observations removed, the average standard deviation in the two practice rounds on the race

179 Greenwald et al., *supra* note 29, at 199–200.

180 *Id.*

181 In the eastern and western samples we reduced the number of trials in the practice rounds (rounds 1, 2, 3, 5, and 6) from twenty to sixteen, so as to save time.

182 Greenwald et al., *supra* note 29, at 213.

183 *Id.* at 214–15.

184 In the race IAT, twenty-nine out of the thirty-three instances in which judges produced latency scores of greater than 10,000 ms on a trial (or 87.9%) occurred during the practice rounds. In the gender IAT, the two instances in which judges exhibited trials that exceeded 10,000 ms occurred in the target round.

IAT was over one second (1064 ms), as compared to 596 ms in the trial rounds. This suggested to us that we ought not to use the practice rounds in the analysis. The practice rounds of the gender IAT were more stable. The standard deviation from the practice rounds (724 ms) was much closer to that of the trial rounds (560 ms). Even though the practice rounds in the gender IAT seemed more stable, for consistency, we dropped these as well. Our measure of the IAT effect for purposes of correlating the IAT scores with judges' decisions was therefore the average difference between the stereotype-congruent round and the stereotype-incongruent round divided by the standard deviation of latencies in both rounds combined. Following Greenwald and his colleagues, we call the measure d' .

Because the latencies that we observed seemed slower than those which have been observed in the Internet study, we assessed the correlation between our two IAT measures and the mean latency. The correlation coefficients between the mean differences and the overall latency were 0.305 on the race IAT and 0.361 on the gender IAT. These correlations are high enough to indicate that our judges have higher IAT scores than other populations simply because they were somewhat slower.¹⁸⁵ The standardized IAT measure using only the trial rounds, however, produced correlations of only 0.046 and 0.002 with the overall mean latencies for the race and sex IATs, respectively. Hence, the d' measure provided a much more reliable measure of the IAT effect than the mean difference.

185 Note that these correlations used all judges, with no exclusions for speed, did not bound the data between 300 and 3000 ms, and did not exclude the first two rounds, as we did for calculating the mean differences.