Implicit Racial Bias Across the Law

Edited by

JUSTIN D. LEVINSON
University of Hawai'i at Mānoa
William S. Richardson School of Law

ROBERT J. SMITH
University of North Carolina, Chapel Hill

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9781107648180

First published 2012

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Implicit racial bias across the law / [edited by] Justin D. Levinson, Roger J. Smith.
p. cm.
Includes bibliographical references.
ISBN 978-1-107-01095-6 (hardback) – ISBN 978-1-107-64818-0 (pbk.)
1. Discrimination in justice administration – United States. 2. Race discrimination – Law and legislation – United States. 3. Bias (Law) – United States. I. Levinson, Justin D. (Justin David), 1974– II. Smith, Roger J. (Roger John), 1948–
KF384.I47 2012
342.7308'73–dc23 2011048919

ISBN 978-1-107-01095-6 Hardback
ISBN 978-1-107-64818-0 Paperback

1

Implicit Racial Bias

A Social Science Overview

Justin D. Levinson, Danielle M. Young, and Laurie A. Rudman

A little after 2:00 a.m. on the first day of 2009, San Francisco Bay Area Rapid Transit (BART) Officer Johannes Mehserle arrived at the Fruitvale BART station after receiving reports of a fight on a train.[1] On arrival, he was directed by another officer to arrest Oscar Grant, who, along with other fight suspects, was sitting on the ground next to the wall of the station. As Mehserle, who was joined by other officers, prepared to arrest Grant, Grant began to stand up, and Mehserle forced him to the ground face first. Another officer stood over Grant and uttered, "Bitch-ass n-."[2] As Mehserle attempted to handcuff Grant, some eyewitnesses testified that Grant resisted by keeping his hands under his torso. Although Grant was lying face down and was physically restrained by another police officer at the time of his alleged resistance, Mehserle removed his department-issued handgun from its holster and shot Grant in the back from point-blank range. Grant died later that morning.

At trial, the jury convicted Mehserle of involuntary manslaughter, but acquitted him of more serious homicide charges that would have treated the killing as intentional. The involuntary manslaughter conviction indicates that the jury likely believed two key pieces of Mehserle's testimony: first, that he thought Grant was reaching for a gun, and second, that he mistook his own gun (which was on the right side of his body and weighed twice as much) for a Tazer (on the left side of his body). How could an officer possibly perceive a mostly compliant, restrained man as a gun-toting threat?

Research on implicit racial bias suggests that, when implicit racial stereotypes are activated, the human mind is capable of major feats, such as turning an

[1] The facts we present are largely based on bail hearing documents and video recordings of the event. For one video recording, *see* thecaliforniabeat, *New Footage of Oscar Grant Shooting*, YouTube (Feb. 24, 2010), http://www.youtube.com/watch?v=KxnFQ_IvOt4.

[2] Philip Matier & Andrew Ross, *BART 'N-word' bombshell waiting to go off*, S.F. CHRONICLE, June 29, 2009, *available at* http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2009/06/28/BA4E18EMPH.DTL&tsp=1.

innocent hand into one reaching for a gun. For more than a decade, researchers have explored how implicit racial bias (as contrasted with self-reported, or explicit, racial bias) contributes to systematic racial discrimination. Implicit racial bias explains the process whereby the human mind automatically and unintentionally reacts to different groups in divergent ways, a process that can have unfortunate consequences. One fundamental aspect of the implicit racial attitudes uncovered by implicit social cognition research is that they frequently differ from people's self-reported (often egalitarian) racial attitudes. Because of the automatic nature of these biases, people are often unaware of them or how they affect their judgments.

In this chapter, we lay the foundation for studying implicit racial bias across the law by examining examples from more than a decade of research on the unconscious and automatic activation of racial stereotypes. This examination includes a discussion of the ease with which racial stereotypes are activated (particularly through the phenomenon known as "priming"), a consideration of how social scientists measure implicit bias (including through the Implicit Association Test and the shooter bias "videogame"), and an exploration of the relationship between implicit bias and real-world behaviors and decision-making. The chapter concludes by describing empirical studies of implicit racial bias in the legal system.

RACIAL PRIMING: THE UNCONSCIOUS ACTIVATION OF STEREOTYPES

Priming is a term imported from cognitive psychology that describes a stimulus that has an effect on an unrelated task. Psychologists have defined it as "the incidental activation of knowledge structures, such as trait concepts and stereotypes, by the current situational context."[3] Simply put, priming studies show how causing someone to think about a particular domain can trigger associative networks related to that domain.[4] Activating these associative networks, which can include stereotypes, can affect people's decision-making and behavior, often without their conscious awareness.

A study conducted by John Bargh and his colleagues provides a simple example of priming outside of the racial context.[5] In the study, which was designed to examine the behavioral effects of activating stereotypes related to the elderly, the researchers primed participants by exposing them to one of two lists of words. Half of the participants unscrambled sentences that contained words designed to subtly prime participants with the category of elderly; thus, the sentences included

[3] John A. Bargh et al., *Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action*, 71 J. PERSONALITY & SOC. PSYCHOL. 230, 230 (1996).

[4] Several of the descriptions we provide of priming and the Implicit Association Test are based, sometimes verbatim, on the descriptions given in Justin Levinson, *Forgotten Racial Equality: Implicit Bias, Decisionmaking, and Misremembering*, 57 DUKE L. J. 345 (2007), and Justin D. Levinson, *Race, Death and the Complicitous Mind*, 58 DEPAUL L. REV. 599 (2009).

[5] Bargh, *supra* note 3.

words such as "wise," "helpless," "wrinkle," and "bingo." The other half of the participants unscrambled sentences using words that contained no particular theme (e.g., "thirsty," "private," and "clean"). The researchers hypothesized that participants who unscrambled sentences containing the elderly stereotyped words would become primed by them. That is, those participants would have activated their usually dormant knowledge of traits and stereotypes of the elderly. Bargh and his colleagues further suggested that simply activating these traits and stereotypes would cause behavioral changes. Specifically, they predicted that participants who were primed for the concept of elderly would walk more slowly down the hallway after believing they had completed the experiment compared to participants who were not primed. This hypothesis was confirmed: primed participants indeed walked down the hallway more slowly than nonprimed participants. Furthermore, participants who were primed with the elderly words were unaware that they had even been exposed to an elderly prime. This example sets the stage for a discussion of the ease with which racial stereotypes may be primed, as well as the important behavioral implications of priming.

Ease of Activation

Racial and ethnic stereotypes can be primed with ease. In one simple and elegant study, participants watched a video in which a research assistant held cue cards containing word fragments.[6] All participants watched identical videos, except for the identity of the research assistant. Half of the participants saw a video in which the research assistant was Asian, and the other half saw a video in which the research assistant was Caucasian. In the video, the assistant held cue cards containing incomplete words, including words that were potentially stereotypic of Asians, such as "RI_E", "POLI_E," "S_ORT," and "S_Y." Participants were asked to generate as many word completions as possible for each card during a fifteen-second period. Results of the study showed that simply seeing an Asian research assistant was enough to activate ethnic stereotypes of Asians. Participants who watched the video with the Asian assistant completed more stereotype words (RICE, POLITE, SHORT, and SHY) than participants who watched the video with the Caucasian assistant.

The cue card study demonstrated that even simple visual cues (seeing an Asian person) can prime a person's racial and ethnic stereotypes. Laurie Rudman and Matthew Lee tested whether auditory, rather than visual, cues could similarly prime participants' racial stereotypes.[7] Participants in their study listened to either rap or pop music for thirteen minutes.[8] As they hypothesized, Rudman and Lee found

[6] Daniel T. Gilbert & J. Gregory Hixon, *The Trouble of Thinking: Activation and Application of Stereotypic Beliefs*, 60 J. PERSONALITY & SOC. PSYCHOL. 509 (1991).

[7] Laurie A. Rudman & Matthew R. Lee, *Implicit and Explicit Consequences of Exposure to Violent and Misogynous Rap Music*, 5 GROUP PROCESSES & INTERGROUP REL. 133, 138–39 (2002).

[8] Participants were led to believe that they were participating in a marketing study.

that rap music indeed activated participants' racial stereotypes, including stereotypes associating African Americans with attributes such as hostile, violent, and dangerous. Furthermore, participants whose racial stereotypes were activated subsequently rated a black (but not a white) person's behavior as less intelligent and more hostile. Finally, participants' self-reported (explicit) prejudice levels did not predict their judgments of a black person, indicating that automatic biases can leak into people's decision-making without their endorsement or awareness. This study demonstrates that even simple auditory primes (such as hearing music) can automatically activate a network of associated implicit racial stereotypes.

Research has also shown that racial stereotypes can be activated by crime-related primes. Jennifer Eberhardt and her colleagues primed some participants with crime-related images,[9] such as a police badge, a fingerprint, and guns, as part of a "dot-probe" task that measures attention by seeing how quickly participants can find a dot on a screen. This dot was presented on a computer screen near the face of either a white male or a black male. Participants who were primed with crime-related images found the dot faster near black faces compared to participants who were not primed, suggesting a high degree of cognitive association between the concepts "crime" and "black male." The researchers found similar results by priming participants with basketball-related words. Participants who had been primed with basketball-related words found the dot faster near black faces than participants who were not primed.

Decision-making and Behavioral Consequences of Priming

Racial stereotypes are not only primed easily in a variety of situations but their activation can also wreak havoc on decision-making and behavior. Keith Payne examined how merely showing participants a photograph of a white or black face for 200 milliseconds could affect the speed at which they could subsequently identify weapons.[10] In the study, participants saw photos of black or white faces followed immediately by photos of objects. The participants' only task, Payne told them, was to quickly identify the objects when they appeared on the screen.[11] Payne also told them that the flashing photographs of faces served only to signal the participant that a photograph of an object was about to appear. The study found that when participants saw photos of black faces immediately before photos of guns, they were significantly faster at identifying the guns. Similarly, when participants saw photos of white faces immediately before photos of tools, they were significantly faster at identifying the tools. Payne's study shows that racial stereotypes can be elicited automatically in a

[9] Jennifer L. Eberhardt et al., *Seeing Black: Race, Crime, and Visual Processing*, 87 J. PERSONALITY & SOC. PSYCHOL. 876 (2004).

[10] B. Keith Payne, *Prejudice and Perception: The Role of Automatic and Controlled Processes in Misperceiving a Weapon*, 81 J. PERSONALITY & SOC. PSYCHOL. 181, 185–86 (2001).

[11] *Id.* The objects consisted of guns and non-gun objects (the non-gun objects were hand tools, such as a socket wrench and an electric drill).

number of milliseconds and that these stereotypes can affect the speed and accuracy of meaningful object classification tasks.

In addition to affecting task performance, priming can affect people's judgments of others' behavior. Patricia Devine had participants watch flashing category words that were associated with African Americans, such as "blacks" and "Negroes," and flashing stereotype words that were associated with African Americans, such as "poor" and "athletic."[12] After the priming was accomplished, participants read a story about a person engaging in ambiguously hostile behaviors – such as demanding money back from a store clerk – and were asked to make judgments about the person engaging in these behaviors. Participants who were primed with more stereotyped words judged the actor's ambiguous behavior as more hostile than participants who were primed with fewer stereotyped words. As Devine summarized, "The automatic activation of the racial stereotype affects the encoding and interpretation of ambiguously hostile behaviors for both high- and low-prejudice subjects."[13] Although traits such as "poor" and "athletic" are unrelated to the trait of "hostile," the stereotype congruence between the primed stereotypes and the trait of hostility made participants more likely to judge a behavior as hostile.

Another study of priming's effects on decision-making was conducted by James Johnson and Sophie Trawalter, who primed participants by playing segments of either a violent or nonviolent rap song.[14] Participants later read supposedly unrelated stories of ambiguous behavior and were asked to make judgments about people in the stories. Those who heard the violent rap music, compared to other participants, judged a black male's aggressive behavior as caused by dispositional factors (for example, a violent personality) rather than situational factors (for example, stress related to a relationship break-up). In addition, participants who heard the violent rap music were more likely to judge a black job applicant as less qualified for a job requiring intelligence. This study shows that racial stereotype primes (here, rap music) can influence seemingly unrelated judgments (here, job qualification) so long as they are broadly related (both are stereotypes of African Americans).

Priming and the Self: Stereotype Threat

Thus far, the priming research we have reviewed shows that in legal contexts, scholars should be on the lookout for racial stereotype priming. Yet, commentators should be concerned with more than the activation of biased decision-making by primed observers. As Bargh's study of priming elderly stereotypes and walking speed shows,

[12] Patricia G. Devine, *Stereotypes and Prejudice: Their Automatic and Controlled Components*, 56 J. PERSONALITY & SOC. PSYCHOL. 5, 9 (1989) (citing J. C. Brigham, *Ethnic Stereotypes*, 76 PSYCHOL. BULL. 15 (1971)).

[13] *Id.* at 11.

[14] James D. Johnson & Sophie Trawalter, *Converging Interracial Consequences of Exposure to Violent Rap Music on Stereotypical Attributions of Blacks*, 36 J. EXPERIMENTAL SOC. PSYCHOL. 233, 239 (2000).

they must also be concerned with the way priming can unconsciously affect the self. Some of the most famous priming experiments have studied the effect of racial, ethnic, and gender stereotypes on students' test-taking performance. In contrast to the foregoing priming studies, the primes in these studies affect a person's performance based on his or her own stereotyped identity. Claude Steele and Joshua Aronson first identified the concept of "stereotype threat" by priming college students in test-taking situations.[15] They primed Caucasian and African American college students by asking them to identify their race just before they took a test. The researchers found that such a simple priming task had profound effects on African American test performance: African American participants took longer to answer questions and achieved lower overall scores than Caucasian participants, but only when they were primed. Thus, Steele and Aronson found that priming a participant's racial identity likely implicated a complex relationship between African American identity and negative stereotypes relating to ability. They called this phenomenon "stereotype threat."

Steele and Aronson also found that stereotype threat could be elicited even by indirectly priming the racial stereotype. In this study, when they told half of the participants that the test results would be used to evaluate performance but did not ask them to identify their race, they found results similar to those obtained when they primed race directly: African American students in the indirect prime condition performed worse than Caucasian students in the same condition, whereas African American and Caucasian students performed similarly in the nonprime condition. This study demonstrates the ease and influence of indirect racial priming. Simply priming a nonexplicit but related stereotype, even without mentioning race, can cause profound results. Although it would not be intuitive to many that using race-neutral concepts can elicit powerful racial stereotypes, social cognition research shows that priming can occur as long as historical, cultural, or popular associations connect the concept with a racial stereotype. In the case of stereotype threat, African American student participants associated the evaluation instruction as implicating negative stereotypes relating to African Americans and intellectual ability.

Follow-up studies of stereotype threat have shown that it can be elicited by using even more indirect primes. Margaret Shih and colleagues used an indirect method of priming student-participants' ethnic identity, but found similarly powerful results.[16] In that study, the researchers asked Asian American female participants to fill out questionnaires before taking a math test. Some questionnaires asked the participants about their roommate and dormitory living situations (this condition was designed to prime gender identity), whereas others asked them about their family, including

[15] Claude M. Steele & Joshua Aronson, *Stereotype Threat and the Intellectual Test Performance of African-Americans*, 69 J. PERSONALITY & SOC. PSYCHOL. 797 (1995).

[16] Margaret Shih et al., *Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance*, 10 PSYCHOL. SCI. 80 (1999).

what languages were spoken at home and how many generations of their family had lived in the United States (this condition was designed to prime ethnic identity). This method of indirect priming significantly affected the participants' test performance. Participants who had their Asian identity primed performed best on the test, whereas participants who had their female identity primed performed worst on the test. Considered together, these studies on stereotype threat show the dangers of subtly priming people's negative stereotypes about the groups to which they belong.

DEFINING THE IMPLICIT – REACTION TIMES, SHOOTER BIAS, AND THE IMPLICIT ASSOCIATION TEST

This chapter has explained how easily racial stereotypes can be primed and how priming can affect decision-making and behavior in troubling ways: for example, in object recognition tasks, when making judgments about others' ambiguous behaviors, and in academic test performance. Each of the reviewed studies shows the dynamic nature of implicit cognitive processes – processes that are important components of human decision-making. This chapter now turns to ways that social scientists measure implicit bias, examining two types of studies that use reaction times to measure implicit bias: shooter bias studies and the Implicit Association Test (IAT).

Quick Trigger Finger: The Shooter Bias

In the aftermath of the killing of Oscar Grant, as had happened with the killing of unarmed black men before him, observers wondered whether Officer Mehserle would have grabbed his gun and pulled the trigger if Grant had been white. Faced with this question after each instance in which police shot unarmed black men, social scientists developed a measure of "shooter bias." Shooter bias studies use custom-made videogames to examine race-based differences in reactions to potentially threatening individuals.[17] The bias can be tested when participants play a videogame that instructs them to shoot perpetrators (who are holding guns) as fast as they can but not to shoot innocent bystanders (who are unarmed but holding a non-gun object, such as a cell phone). The "shooter bias" refers to the consistent results of these studies: participants tend to shoot black perpetrators more quickly and more frequently than white perpetrators and conversely decide not to shoot white bystanders more quickly and frequently than black bystanders.

Once shooter bias became an established phenomenon in study participants who were not police officers, researchers wondered whether police officers would also display the bias. After all, unlike civilians, police officers receive extensive handgun

[17] Joshua Correll et al., *The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals*, 83 J. PERSONALITY & SOC. PSYCHOL. 1314 (2002).

training and exercise visual discrimination tasks (such as detecting a gun) as a regular part of their job. Perhaps, then, police officers could resist or overcome the bias shown by other citizens. Joshua Correll and his colleagues tested this question by using both a community sample and a sample of police officers.[18] They found that, although police officers were generally faster and more accurate than the community sample, their reaction times followed the same pattern as that of community members: police officers were faster to "shoot" armed black perpetrators than armed white perpetrators and took longer to make "don't shoot" decisions for unarmed black targets than for unarmed white targets.

Researchers have investigated the cognitive roots of shooter bias, hypothesizing that it may manifest in brain processes that moderate responses to fear. To that end, Correll and his colleagues looked at fluctuations in participants' electrical brain activity (known as "event-related brain potentials") while the participants played the shooter bias videogame.[19] Measuring event-related brain potentials can identify when people detect threats and when they have a desire to control a behavioral response. The results of the study showed that, as participants played the videogame, racial discrepancies manifested in the electrical activity of their brains. That is, participants' brain activity showed more threat-related brain activity when presented with black actors than white actors (even for black actors without guns) and more control response activity for white actors than black actors. These brain responses correlated with the participants' performance – the more biased brain activity they displayed, the more shooter bias they exhibited.

The Implicit Association Test

Shooter bias studies are not the only social science measure that uses reaction times and accuracy rates to measure potential racial bias. Within legal discourse, the most frequently discussed measure of implicit social cognition is the Implicit Association Test (IAT). The IAT pairs an attitude object (such as a racial group) with an evaluative dimension (good or bad) and tests how response accuracy and speed indicate implicit and automatic attitudes and stereotypes. Participants sit at a computer and are asked to pair an attitude object (for example, black or white, man or woman, fat or thin) with either an evaluative dimension (for example, good or bad) or an attribute dimension (for example, home or career, science or arts) by pressing a response key as quickly as they can. For instance, in one task,

[18] Joshua Correll et al., *Across the Thin Blue Line: Police Officers and Racial Bias in the Decision to Shoot*, 92 J. PERSONALITY & SOC. PSYCHOL. 1006 (2007). The study looked at two measures of shooter bias: response times and accuracy. This chapter reports the results for the reaction time study in the text. The study of accuracy demonstrated that police officers were generally more accurate in their decisions to shoot than the community sample.

[19] Joshua Correll et al., *Event-Related Potentials and the Decision to Shoot: The Role of Threat Perception and Cognitive Control*, 42 J. EXPERIMENTAL SOC. PSYCHOL. 120, 122 (2006).

participants are told to quickly pair pictures of African American faces with positive words from the evaluative dimension. In a second task, participants are obliged to pair African American faces with negative words. The difference in the speed at which the participants can perform the two tasks is interpreted as the strength of the attitude (or, in the case of attributes, the strength of the stereotype). For example, if participants perform the first task faster than the second task, they are showing implicitly positive attitudes toward blacks. Similarly, if they are faster to perform tasks that oblige categorizing women with home than tasks that oblige categorizing women with career, they are showing implicit sex stereotyping.

Nilanjana Dasgupta and Anthony Greenwald succinctly summarize the science underlying the IAT: "When highly associated targets and attributes share the same response key, participants tend to classify them quickly and easily, whereas when weakly associated targets and attributes share the same response key, participants tend to classify them more slowly and with greater difficulty."[20] Laurie Rudman and Richard Ashmore add, "The ingeniously simple concept underlying the IAT is that tasks are performed well when they rely on well-practiced associations between objects and attributes."[21]

Scores of studies have found that people harbor implicit associations that are biased against stereotyped group members.[22] According to Brian Nosek and his colleagues, who reviewed hundreds of thousands of IATs taken on the web and elsewhere, the IAT has consistently shown that a majority of test takers exhibit implicit racial bias – and other nonracial biases – on a variety of measures.[23] For example, 68 percent of participants demonstrated an implicit preference for "white people" versus "black people" (or "light skin" versus "dark skin"), 75 percent of participants showed an implicit preference for "abled people" versus "disabled people," and 69 percent of participants showed an implicit preference for "thin people" versus "fat people." Similar to other community members, law students have also been shown to harbor implicit biases, such as showing an implicit association between men and judges (and women and paralegals).[24]

One particularly interesting characteristic of IAT results, as well as the results of other implicit measures, is that they frequently diverge from self-reported (explicit) racial attitudes. That is, people who display strong implicit biases are often not

[20] Nilanjana Dasgupta & Anthony G. Greenwald, *On the Malleability of Automatic Attitudes: Combating Automatic Prejudice with Images of Admired and Disliked Individuals*, 81 J. PERSONALITY & SOC. PSYCHOL. 800 (2001).

[21] Laurie A. Rudman & Richard D. Ashmore, *Discrimination and the Implicit Association Test*, 10 GROUP PROCESSES & INTERGROUP REL. 359 (2007).

[22] *See* Anthony Greenwald et al., *Understanding and Using the Implicit Association Test: III. Meta Analysis of Predictive Validity*, J. PERSONALITY & SOC. PSYCHOL. (2009).

[23] Brian Nosek et al., *Pervasiveness and Correlates of Implicit Attitudes and Stereotypes*, 18 EUR. REV. SOC. PSYCHOL. 36 (2008).

[24] Justin D. Levinson & Danielle Young, *Implicit Gender Bias in the Legal Profession: An Empirical Study*, 18 DUKE J. GENDER L. & POL'Y 1 (2010).

the same people who claim to have strong explicit biases.[25] According to Devine, "[e]ven those who consciously renounce prejudice have been shown to have implicit or automatic biases that conflict with their nonprejudiced values."[26] For example, in the context of shooter bias, explicit measures of racial preferences do not correlate with results for the videogame. That is, people who exhibit greater amounts of shooter bias are not necessarily the same ones who endorsed more racially unequal preferences.[27] A related and somewhat surprising finding in IAT research has been that members of bias-affected groups, who at least self-report favoring their own group, sometimes harbor implicit biases against their own group. For example, both male and female law students have been shown to display implicit gender stereotypes that are sometimes negative toward women.[28] Similarly, Nosek and his colleagues reported that older people show an implicit preference for young over old.[29]

Despite its consistent results, the IAT has not been without critique.[30] One question regarding the IAT is whether it measures something entirely unconscious or only partially unconscious. Russell Fazio and Michael Olson explored this question and argued that it is difficult, if not impossible, to know if these associations are in fact completely unknown to the participant. In fact, in many nonsensitive domains, the IAT has been shown to correlate well with explicit measures (such as voter intention).[31] It is important to acknowledge that, although the entirely unconscious nature of the attitudes tested by the IAT may be legitimately questioned, support for the automaticity of the association is unquestioned because results are based on swift reaction times. This automaticity assertion is bolstered by research demonstrating

[25] See Patricia G. Devine, *Implicit Prejudice and Stereotyping: How Automatic Are They? Introduction to the Special Section*, 81 J. PERSONALITY & SOC. PSYCHOL. 757, 757 (2001).

[26] *Id.*

[27] *Id.* Not all studies show no relationship between implicit and explicit attitudes. Some studies reveal at least a weak correlation between the two. *See, e.g.*, Russell H. Fazio & Michael A. Olson, *Implicit Measures in Social Cognition Research: Their Meanings and Use*, 54 ANN. REV. PSYCHOL. 297, 304 (2003) (observing that there is "no simple answer" to the issue of whether and how implicit and explicit attitudes are related, but nonetheless suggesting that both are predictive of behavior in different ways); Wilhelm Hofmann et al., *A Meta-Analysis on the Correlation Between the Implicit Association Test and Explicit Self-Report Measures*, 31 PERSONALITY & SOC. PSYCHOL. BULL. 1369, 1382 (2005) (finding a relationship between implicit and explicit attitudes).

[28] Levinson & Young, *Implicit Gender Bias*, *supra* note 24.

[29] Nosek et al., *supra* note 23.

[30] Within legal scholarship, a few scholars have cautioned against embracing the results of the IAT as a measure of bias to be considered in lawmaking. *See, e.g.* Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 OHIO ST. L.J. 1023 (2006). Most psychologists and legal scholars, however, have argued that the IAT has been sufficiently validated as a measure. *See, e.g.*, Samuel R. Bagenstos, *Implicit Bias, "Science," and Antidiscrimination Law*, 1 HARV. L. & POL'Y REV. 477 (2007); John T. Jost et al., *The Existence of Implicit Bias Is Beyond Reasonable Doubt: A Refutation of Ideological and Methodological Objections and Executive Summary of Ten Studies that No Manager Should Ignore*, 29 RES. IN ORGANIZATIONAL BEHAV. 39 (2009).

[31] *See* Greenwald et al., *supra* note 22.

that as people age, their decreased ability to inhibit biased responses manifests in IAT scores.[32]

The IAT Predicts Real-World Behaviors

Social scientists have also been captivated by the question of whether the IAT has predictive validity, that is, the ability to predict real-life behaviors. One of the most pressing questions surrounding measurement in psychology is whether a measurement means anything in the real world. If a person possesses an implicit dislike of snakes, but shows no signs of bias against snakes in the real world, what is that person's implicit bias actually measuring? And is that measure useful? Predictive validity research helps answer such questions. It establishes the validity of an implicit measure (such as the IAT) by determining its relationship with a reasonable real-world measure. The IAT has been shown to predict discriminatory decision-making and behavior in a broad range of ways. Here, we review three particularly compelling studies of the IAT's predictive validity and then discuss a meta-analysis that was conducted on more than one hundred IAT studies.

Rudman and Ashmore tested whether the IAT predicted economic discrimination.[33] Student-participants first took a series of IATs, including those testing negative stereotypes related to Jews and Asians. On a separate occasion, the same participants completed a survey designed to test economic discrimination. Participants were told that their input was needed in determining how to administer a mandatory 20 percent budget cut to university student organizations. They were then provided a list of current student organizations along with funding levels and were asked to allocate the new, reduced budget across the various groups. The researchers then compared the student-participants' IAT scores with their recommended budget cuts and found that scores on the stereotype IAT predicted economic discrimination. Specifically, "people who associated minority group members with negative attributes and majority group members with positive attributes were also likely to recommend budget cuts for the target minority group's student organization."[34] Rudman and Ashmore's study demonstrates a meaningful connection between implicit racial bias and economic inequality.

In a study that linked implicit racial bias to inequality in the provision of health care services, Alexander Green and his colleagues tested whether physicians held implicit racial bias against African Americans and whether this bias predicted their decisions to treat patients.[35] Nearly 300 emergency room and internal medicine

[32] Karen Gonsalkorale et al., *Aging and Prejudice: Diminished Regulation of Automatic Race Bias Among Older Adults*, 45 J. EXPERIMENTAL SOC. PSYCHOL. 410 (2009).

[33] Rudman & Ashmore, *supra* note 21.

[34] *Id.* at 367.

[35] 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 (2007).

physicians in Boston and Atlanta participated in the study. They were presented with a vignette in which a patient, who was described as either black or white, arrives at an emergency room suffering from acute coronary syndrome. The doctors were asked to recommend a course of treatment for the patient and were then asked to complete three IATs testing their implicit racial biases. The study showed that these physicians not only implicitly preferred white patients to black patients but also that their implicit racial biases predicted whether or not they would recommend thrombolysis (clot-busting) treatment to a white or black patient suffering from myocardial infarction. The more the doctors implicitly preferred the white patients, the more likely they were to recommend thrombolysis treatment to white but not black patients. No similar predictive validity was found by asking doctors about their explicit racial preferences. On average, the physicians self-reported no racial preferences at all.

Implicit racial bias has also been shown to predict employment discrimination. Dan-Olof Rooth examined whether human resources officers at corporations harbored implicit bias that affected their choices of which candidates to interview for vacant positions.[36] In the first stage of his study, Rooth responded to more than 1,500 job postings in Sweden for a variety of jobs, ranging from computer professionals to motor vehicle operators. For each job posting, Rooth sent two equal resumes, with the only difference being whether the applicant's name appeared to be Swedish or Arabic/Muslim. He then measured whether the fictitious candidates were summoned for interviews. After this first interview stage was complete, Rooth tracked down the human resources officers responsible for the hiring decisions and invited them to participate in his study (these participants were unaware of the bias-related purpose of the study). Rooth provided an IAT testing implicit racial stereotypes of Swedes and Arab/Muslims and evaluated whether it predicted the human resources officers' previous decision of whether to interview the applicants. The study found that the human resource officers' implicit racial stereotypes of Arabs predicted whether they would call Arab job candidates for interviews. Human resources officers who held implicit stereotypes relating to Arabs were less likely to interview candidates with Arab-sounding names.

Taken individually, these studies on the IAT as a predictor are cause for concern. Yet a recent meta-analysis confirms the predictive validity of the IAT generally, particularly when it is employed in socially sensitive domains such as race. Greenwald and colleagues analyzed 122 studies that mapped IAT scores onto various predictors, such as behaviors, judgments, or physiological measures.[37] They included in their analyses a comprehensive list of IAT studies (going beyond those on race) that

[36] Dan-Olof Rooth, *Automatic Associations and Discrimination in Hiring: Real World Evidence*, 17 LABOUR ECON. 523 (2010).

[37] Greenwald et al., *supra* note 22.

tested a range of implicit attitudes, stereotypes, self-concepts, and self-esteem.[38] The researchers coded each study on a number of items including social sensitivity (a study of implicit racial bias, for example, was coded as being highly socially sensitive). Results showed that the IAT's relationship to predictive validity measures varied, ranging from low to high. Overall, explicit measures had a larger relationship with the response criterion. However, when dealing with socially sensitive issues such as race, the relationship between explicit measures and the response criterion was diminished, whereas the relationship between the IAT and response criterion was not. In fact, when looking specifically at interracial (or other intergroup) topics, IATs were more accurate than explicit measures in predicting behaviors, judgments, and physiological responses. This result confirms that implicit biases, particularly in the context of race, are meaningful. It would be advisable, then, if scholars concerned with racial bias in the legal system systematically examined implicit racial bias across the law.

IMPLICIT BIAS RESEARCH IN THE LEGAL SETTING

Although the vast majority of empirical studies on implicit bias have been conducted outside of the legal setting, a limited number of studies have used priming procedures, IATs, or both to measure bias in the law. As the following review demonstrates, significant progress has been made in empirically examining implicit racial bias in a few narrow domains, but most legal areas have generally been overlooked.

The first empirical use of the IAT in the legal setting occurred when Theodore Eisenberg and Sheri Lynn Johnson tested whether capital defense attorneys harbor implicit racial bias.[39] The researchers found that the defense attorney participants, a group one might expect to resist bias, in fact harbored strong implicit bias against African Americans. Eisenberg and Johnson, however, did not test whether the defense attorneys' implicit bias predicted anything about their behavior or decisions. Nonetheless, this study documented implicit bias among a particularly noteworthy participant population, and it opened the door for future research.

An early empirical examination of priming in the legal setting was conducted by Justin Levinson in 2005.[40] Levinson hypothesized that simply placing citizens on juries activates implicit and explicit knowledge structures that change the way people make decisions. He provided study participants with the facts of a crime and asked half of them to imagine that they were jurors in the criminal trial. The other half of the

[38] The resulting studies included a range of areas of research that dealt with topics from smoking to racial inequality.

[39] *See* Theodore Eisenberg & Sheri Lynn Johnson, *Implicit Racial Attitudes of Death Penalty Lawyers*, 53 DEPAUL L. REV. 1539, 1542 (2004).

[40] Justin D. Levinson, *Suppressing the Expression of Community Values in Jurors: How "Legal Priming" Systematically Alters the Way People Think*, 73 U. CIN. L. REV. 1059 (2005).

participants were informed that they were simply reading newspaper accounts of the crime. The results of the study showed that mock juror participants were significantly harsher in making judgments of criminal intentionality than participants making lay judgments about the same facts. In addition, when the defendant was portrayed as an outgroup member, participants in the legal prime condition appeared to become even harsher in their decisions.

In a law-focused study that employed both the IAT and priming, Jeffrey Rachlinski, Sheri Lynn Johnson, and their colleagues examined whether judges harbored implicit bias, and tested whether the IAT could predict race-based judicial decision-making.[41] Judges completed a black/white IAT and subsequently decided hypothetical court scenarios in which the race of the legal actor was subliminally primed. The researchers found first that judges displayed an implicit preference for white over black, and second, that the IAT predicted responses in the judgment tasks in which the race of the legal actor had been primed subliminally. For example, when the defendant was subliminally primed to be black, judges who scored in a pro-white direction on the IAT handed down harsher sentences.

Levinson and Danielle Young tested how priming mock jurors with the image of a dark-skinned perpetrator might alter judgments about the probative value of evidence.[42] Levinson and Young provided all jurors with the story of an armed robbery. After reading the story, participants were shown five crime scene photos for four seconds each. All participants saw the same five photos, except that in the third photo, half of the participants saw a dark-skinned perpetrator and the other half saw a lighter skinned perpetrator.[43] The researchers then presented participants with various pieces of evidence that were described as trial testimony and instructed them to rate each piece of evidence based on its probative value. As hypothesized, Levinson and Young found that participants who had seen a photo of a darker skinned perpetrator were more likely to evaluate the evidence as tending to indicate guilt.

Levinson, Huajian Cai, and Young also created an IAT specifically for the criminal law context.[44] Similar to Rachlinski and colleagues' study of judges, Levinson and his colleagues examined the predictive validity of the IAT. In this project, however, the researchers designed a Guilty/ Not Guilty IAT that they expected might be a more meaningful measure in the legal context. Levinson, Cai, and Young, as predicted, found that participants held a strong implicit association between black and guilty

[41] *See* Jeffrey J. Rachlinski et al., *Does Unconscious Bias Affect Trial Judges?*, 84 NOTRE DAME L. REV. 1195 (2009).

[42] Justin D. Levinson & Danielle Young, *Different Shades of Bias: Skin Tone, Implicit Racial Bias, and Judgments of Ambiguous Evidence*, 112 W. VA. L. REV. 307 (2010).

[43] None of the other photos showed the perpetrator.

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

compared to white and guilty. In addition, they found that the IAT scores predicted participants' evidence judgments.[45]

In another project that employed an IAT tailored specifically for the legal setting, Jerry Kang and his colleagues created an IAT designed to test whether people rely on implicit ethnic biases when evaluating the performance of litigators.[46] Specifically, the researchers tested implicit bias and related evaluations for Asian male litigators compared to white male litigators. The researchers predicted that participants would implicitly associate white males with characteristics frequently associated with successful litigators relative to Asian males, who would be more likely to be associated with characteristics frequently associated with successful scientists. As the results of the study showed, participants did in fact implicitly associate white males with successful litigators compared to Asian males. In addition, these implicit associations showed predictive validity. The IAT scores predicted participants' judgments of white and Asian litigators' performance in a mock trial.

A study by Young, Levinson, and Scott Sinnett used priming to follow up Levinson and colleagues' Guilty/ Not Guilty IAT, which raised the issue that the presumption of innocence may have a different implicit meaning for white and black defendants.[47] The researchers hypothesized that the presumption of innocence itself would actually prime participants to think about guilty African Americans, a counterintuitive result that might call into question the racial fairness of the presumption of innocence. Participants watched a video recording of a U.S. District Court judge reading a series of jury instructions. Half of the participants received instructions that included a presumption of innocence instruction, and the other half received an alternative matched-length instruction. After receiving the jury instruction, which the researchers hypothesized would act as a racial prime, participants completed a computer-based dot-probe task to assess if participants' attention was primed for black faces. Participants who were given the presumption of innocence instruction were faster to find a dot when it appeared on the same side of the screen as black faces than when it appeared on the same side of the screen as white faces. Participants who did not receive the presumption of innocence instructions displayed a similar response time for both white and black faces. These results demonstrate that presumption of innocence instructions prime attention for black faces.

[45] *Id.* at 206. The Guilty-Not Guilty IAT scores predicted overall judgments of evidence, but did not predict these judgments based on the skin tone of the perpetrator. Later analysis showed that Pleasant-Unpleasant IAT scores predicted judgments based on the skin tone of the perpetrator.

[46] Jerry Kang et al., *Are Ideal Litigators White? Measuring the Myth of Colorblindness*, 7 J. EMPIRICAL LEG. STUD. 886 (2010). Although we do not describe it in the text, another 2010 study employed the IAT to test whether law students possess implicit gender biases related to women in the legal profession. *See* Levinson & Young, *supra* note 24 (finding, for example, that law students associate men with judges and women with paralegals).

[47] Danielle Young, Justin D. Levinson, & Scott Sinnett, Presumption of Innocence: Biasing Racial Cues (2011) (unpublished manuscript) (on file with authors).

The foregoing studies show the progress made in empirically testing implicit bias in the legal domain. Yet, the summaries also underscore the limited nature of these studies. Future research should continue to empirically investigate implicit racial bias in the legal setting.

CONCLUSION

In light of the evidence linking implicit racial bias to a variety of discriminatory outcomes, legal scholars and empiricists must consider deeply the various ways in which implicit bias may affect all areas of the law in which disparities appear. The killing of Oscar Grant, which reopened old wounds for many Americans clinging to the prospect of a future with equal justice, serves as a stark reminder of the powerful role that racial stereotypes can play, even in a society that espouses racial equality. Although significant research has begun to pave the way for progress in the legal system, researchers must continue paving a path to justice.