

Guilty By Implicit Racial Bias: The Guilty/Not Guilty Implicit Association Test

Justin D. Levinson,^{*} Huajian Cai,^{**} and Danielle Young^{***}

I. INTRODUCTION

Legal scholarship on racial discrimination has turned to the science of implicit social cognition to explain how the human mind automatically manifests biases against disfavored social groups.¹ Much of this discourse on implicit bias focuses on the potential for massive, but hard to detect discrimination in the employment context.² Yet, other legal domains where implicit racial bias may lead to persistent racial inequalities remain underexplored, most notably in criminal law. Specifically, a crucial question still needs to be answered: do implicit biases affect jury guilty/not guilty verdicts in racially biased ways?

Despite the broad incorporation of social science knowledge into legal discourse, a critical chasm continues to deter legal scholarship from fully achieving the social cognition-informed perspective it craves. Namely, legal scholarship on implicit bias lacks law-focused science.³ Legal analysts have implicitly assumed

Copyright © 2010 by Justin D. Levinson, Huajian Cai, and Danielle Young

^{*} Associate Professor of Law, University of Hawai'i. The authors would like to thank Susan Serrano, Dina Shek, Kapua Sproat, and Elizabeth Page-Gould for their contributions to earlier stages of this project. Sara Ayabe provided outstanding research assistance. Dean Aviam Soifer supplied generous summer research support to the first author. Please address correspondence to Justin Levinson, justinl@hawaii.edu, or Huajian Cai, caihj@psych.ac.cn.

^{**} Professor, Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences.

^{***} Department of Psychology, University of Hawai'i.

¹ See, e.g., Jerry Kang, *Trojan Horses of Race*, 118 HARV. L. REV. 1489, 1497–1539 (2005); Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161 (1995); Justin D. Levinson, *Forgotten Racial Equality: Implicit Bias, Decisionmaking, and Misremembering*, 57 DUKE L.J. 345 (2007) [hereinafter *Forgotten Racial Equality*]; Antony Page, *Batson's Blind-Spot: Unconscious Stereotyping and the Peremptory Challenge*, 85 B.U. L. REV. 155 (2005).

² See Samuel R. Bagenstos, *The Structural Turn and the Limits of Antidiscrimination Law*, 94 CALIF. L. REV. 1 (2006) [hereinafter *The Structural Turn*]; Melissa Hart, *Subjective Decisionmaking and Unconscious Discrimination*, 56 ALA. L. REV. 741 (2005); Krieger, *supra* note 1; Ann C. McGinley, ¡Viva La Evolucion!: *Recognizing Unconscious Motive in Title VII*, 9 CORNELL J.L. & PUB. POL'Y 415 (2000); Deana A. Pollard, *Unconscious Bias and Self-Critical Analysis: The Case for a Qualified Evidentiary Equal Employment Opportunity Privilege*, 74 WASH. L. REV. 913 (1999); Audrey J. Lee, Note, *Unconscious Bias Theory in Employment Discrimination Litigation*, 40 HARV. C.R.-C.L. L. REV. 481 (2005).

³ There have been a few empirical studies of implicit bias in the legal setting. See *infra* Section II-C for an overview of this empirical legal scholarship.

that existing social cognition measures, many of which are carefully developed and rigorously tested (but not developed with the law in mind), are the only options for theory development in the legal context. These tests have been groundbreaking in social psychological scholarship and their introduction into legal scholarship has addressed the need for the law to possess an understanding of the human mind. Yet, the still emerging legal model of the human mind has failed to develop new empirical tests that measure how implicit cognitive processes function not just in society in general, but specifically in legally relevant contexts such as jury decision-making.

Here is one example: a frequently cited psychological measure of implicit bias, the Implicit Association Test (“IAT”), examines people’s implicit associations by measuring response speed in a computerized test.⁴ In one of the most famous IATs, study participants are asked to pair together words representing attitudes (Good and Bad) and photos depicting target group members (Black and White) as fast as they can.⁵ The results of these studies show that, when measuring response times and error rates, the vast majority of people are faster to pair together Good with White and Bad with Black.⁶ These results are considered to be indicative of implicit bias, and are eye-opening when considered in the legal context.⁷ Yet might these studies do even more to examine implicit bias in the legal system? For example, why should legal scholars be satisfied to rely on psychological research relating to implicit racial attitudes of “good” and “bad,” (and then engage in heated debate about what it really means in the legal context⁸) when it is possible to

⁴ Researchers call this “response latency.” See, e.g., Anthony G. Greenwald et al., *Measuring Individual Differences in Implicit Cognition: The Implicit Association Test*, 74 J. PERSONALITY & SOC. PSYCHOL. 1464, 1467 (1998) [hereinafter *Measuring Individual Differences*]; Mahzarin R. Banaji, *Implicit Attitudes Can Be Measured*, in THE NATURE OF REMEMBERING: ESSAYS IN HONOR OF ROBERT G. CROWDER 123 (Henry L. Roediger III et al. eds., 2001).

⁵ See Banaji, *supra* note 4, at 123, 136; Greenwald et al., *supra* note 4, at 1465; Jeffrey J. Rachlinski et al., *Does Unconscious Bias Affect Trial Judges?*, 84 NOTRE DAME L. REV. 1195, 1198–99 (2009) [hereinafter *Trial Judges*].

⁶ See Brian A. Nosek et al., *Harvesting Implicit Group Attitudes and Beliefs from a Demonstration Website*, 6 GROUP DYNAMICS: THEORY RES. & PRAC. 101, 105–06 (2002).

⁷ See Kang, *supra* note 1, at 1493 (calling implicit social cognition findings “stunning”).

⁸ Although legal debates involving the IAT have mostly focused on attitude IATs, these debates are both broad in scope and sophisticated, and focus on everything from the meaning of reaction times (and the particular methods of scoring those reaction times) to issues of predictive validity (the issue of whether the IAT predicts anything meaningful). See Samuel R. Bagenstos, *Implicit Bias, “Science,” and Antidiscrimination Law*, 1 HARV. L. & POL’Y REV. 477 (2007) [hereinafter “*Science*” and *Antidiscrimination Law*]; Adam Benforado & Jon Hanson, *Legal Academic Backlash: The Response of Legal Theorists to Situationist Insights*, 57 EMORY L.J. 1087 (2008); Gregory Mitchell & Philip E. Tetlock, *Facts Do Matter: A Reply to Bagenstos*, 37 HOFSTRA L. REV. 737 (2009) [hereinafter *Facts Do Matter*]; Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 OHIO ST. L.J. 1023, 1028, 1032–33 (2006) [hereinafter *Perils of Mindreading*] (arguing that some legal scholars tend to propose legal solutions to the IAT without investigating alternative causes for the IAT’s results).

specifically test implicit associations of well known legally meaningful constructs, such as “guilty” and “not guilty?”⁹

To address the lack of legally-focused empirical studies exploring implicit bias, we developed a new IAT: the Black/White, Guilty/Not Guilty IAT (“Guilty/Not Guilty IAT”). We designed this IAT to examine whether people hold implicit associations between African Americans¹⁰ and criminal guilt, a finding that would call into question criminal law’s presumption of innocence and evoke larger questions of racial justice. Although the debate over racial disparities in the criminal justice system has been raging for decades,¹¹ scholars have rarely adapted social cognition methodology to examine specifically the role of race in criminal law decision-making.¹² We therefore created and developed the Guilty/Not Guilty IAT, and predicted that people implicitly associate Black and Guilty compared to White and Guilty. Because it is important not just to test implicit associations

⁹ Like the debates on the Good-Bad attitude IAT, scholars should critique all law-focused social science.

¹⁰ Because the IAT measures implicit associations related to photos of Black males, we cannot always know specifically that IAT results are due to stereotypes of African Americans rather than stereotypes of other Black males. However, social cognition researchers consistently find that using photos of Black males triggers stereotypes of African American males. *See, e.g.*, B. Keith Payne, *Prejudice and Perception: The Role of Automatic and Controlled Processes in Misperceiving a Weapon*, 81 J. PERSONALITY & SOC. PSYCHOL. 181, 187 (2001) (finding that showing participants Black faces for 200 milliseconds acted to trigger racial stereotypes associated with African Americans). To confirm that we were testing stereotypes of African American men, in our empirical study we specifically referred to the target group as “African American.”

¹¹ For more on racial disparities in the criminal justice system, see, for example, RANDALL KENNEDY, *RACE, CRIME, AND THE LAW* (1997); FROM LYNCH MOBS TO THE KILLING STATE: RACE AND THE DEATH PENALTY IN AMERICA (Charles J. Ogletree, Jr., & Austin Sarat eds., 2006); Scott Phillips, *Racial Disparities in the Capital of Capital Punishment*, 45 HOUS. L. REV. 807, 811–12 (2008).

¹² There are a few studies where scholars have employed social cognition methodologies in the legal setting. *See* Levinson, *Forgotten Racial Equality*, *supra* note 1, at 354; Rachlinski et al., *Trial Judges*, *supra* note 5; Samuel R. Sommers, *On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations*, 90 J. PERSONALITY & SOC. PSYCHOL. 597 (2006); Samuel R. Sommers & Michael I. Norton, *Race-Based Judgments, Race-Neutral Justifications: Experimental Examination of Peremptory Use and the Batson Challenge Procedure*, 31 LAW & HUM. BEHAV. 261 (2007). For a discussion of some of these studies, see *infra* notes 42–55 and accompanying text. Interestingly, outside of the context of race, quite a few legal scholars have relied on social cognition research and cognitive psychology more broadly. One such area is in developing models of jury decision-making. *See generally* Nancy Pennington & Reid Hastie, *A Cognitive Theory of Juror Decision Making: The Story Model*, 13 CARDOZO L. REV. 519 (1991); Nancy Pennington & Reid Hastie, *Explaining the Evidence: Tests of the Story Model for Juror Decision Making*, 62 J. PERSONALITY & SOC. PSYCHOL. 189 (1992); Dan Simon, *A Third View of the Black Box: Cognitive Coherence in Legal Decision Making*, 71 U. CHI. L. REV. 511 (2004). Another widely researched area that usually does not consider race, but frequently incorporates empirical social cognition work, is behavioral law and economics. *See, e.g.*, BEHAVIORAL LAW & ECONOMICS (Cass Sunstein ed., 2000); Christine Jolls et al., *A Behavioral Approach to Law and Economics*, 50 STAN. L. REV. 1471 (1998).

themselves, but to investigate whether they predict meaningful behaviors,¹³ we also tested whether responses on the Guilty/Not Guilty IAT predict the way mock jurors evaluate ambiguous trial evidence. The results of our study confirmed our hypotheses: study participants held strong associations between Black and Guilty, relative to White and Guilty, and these implicit associations predicted the way mock jurors evaluated ambiguous evidence. Furthermore, we compared our measure to a frequently administered IAT that tests positive and negative attitudes towards race, the Pleasant/Unpleasant IAT, and found that the Guilty/Not Guilty IAT and the Pleasant/Unpleasant IAT functioned differently, a result that demonstrates the uniqueness of the Guilty/Not Guilty measure.

This Article introduces the Guilty/Not Guilty IAT, details the empirical study we conducted, and argues for the need to increase collaborations to employ social cognition methods to test legal hypotheses. Section II presents an overview of IAT research in the legal context, and notes the limited number of empirical studies that have been employed. Section III sets the stage for our empirical study, first by reviewing the science behind the IAT, and second, by contextualizing the meaning of the Guilty/Not Guilty IAT within the doctrine of the presumption of innocence. Section IV details the empirical study we conducted. The study tested implicit associations within an important legal domain and examined whether these implicit associations matter in legal decision-making. Results of the study showed that participants held implicit associations between Black and Guilty compared to White and Guilty, and that these implicit associations predicted mock-juror evaluations of ambiguous evidence. Section V briefly discusses the implications of the study, and calls for increased empirical collaborations. Section VI concludes.

II. IMPLICIT BIAS AND THE IMPLICIT ASSOCIATION TEST

Legal scholarship on implicit bias has emerged rapidly since 2005.¹⁴ By engaging in a science-based dialogue and by endeavoring to understand the

¹³ See, e.g., Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity*, 97 J. PERSONALITY & SOC. PSYCHOL. 17 (2009) (showing that the IAT predicts behaviors in many circumstances); Rachlinski et al., *Trial Judges*, *supra* note 5. See also Hart Blanton et al., *Strong Claims and Weak Evidence: Reassessing the Predictive Validity of the IAT*, 94 J. APPLIED PSYCHOL. 567 (2009) (disputing the results of prior articles claiming that the race IAT predicts behaviors); Allen R. McConnell & Jill M. Leibold, *Weak Criticisms and Selective Evidence: Reply to Blanton et al. (2009)*, 94 J. APPLIED PSYCHOL. 583 (2009) (responding to Blanton et al.'s critique); Jonathan C. Ziegert & Paul J. Hanges, *Strong Rebuttal for Weak Criticisms: Reply to Blanton et al. (2009)*, 94 J. APPLIED PSYCHOL. 590 (2009) (criticizing Blanton et al.'s critique of their prior data); Hart Blanton et al., *Transparency Should Trump Trust: Rejoinder to McConnell and Leibold (2009) and Ziegert and Hanges (2009)*, 94 J. APPLIED PSYCHOL. 598 (2009) (maintaining their skepticism about whether the IAT reliably predicts discriminatory behavior).

¹⁴ Work on what many legal scholars have called "unconscious bias" was introduced conceptually in the 1980s by Charles Lawrence as part of an exploration of anti-discrimination law.

complexities of the human mind, this scholarship has opened up new ways of understanding societal inequality. This section sets the stage for our empirical study of implicit racial bias by examining one of the most compelling measures of implicit bias, the IAT.¹⁵ The section first explains the IAT itself. It then reviews legal scholarship that specifically discusses the IAT. Finally, it considers the few instances of legal scholarship employing empirical methods to run the IAT in legal context.

A. *The Science of the IAT*

The IAT measures implicit cognitions in a simple and compelling way. It asks participants to categorize information as quickly as possible, and then calculates a participant's reaction time (in milliseconds) and accuracy in completing the categorization task.¹⁶ The wisdom behind the IAT holds that statistically significant speed and accuracy-based differences in a person's ability to categorize different types of information reflect something meaningful in that person's automatic cognitive processes.

See Charles R. Lawrence III, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317, 331–36 (1987). See also Krieger, *supra* note 1 (introducing social cognition research on unconscious bias to the employment discrimination context). More contemporary work has evolved rapidly since Jerry Kang's 2005 examination of implicit social cognition in the context of Federal Communications Commission policy. See Kang, *supra* note 1. See also Justin D. Levinson, *Race, Death, and the Complicitous Mind*, 58 DEPAUL L. REV. 599 (2009) [hereinafter *The Complicitous Mind*]; Levinson, *Forgotten Racial Equality*, *supra* note 1; Page, *supra* note 1. A few law-based publications pre-2005 do mention the IAT, but the intense theoretical debates over the IAT heated up with Kang's article. For articles considering the IAT prior to 2005, see, for example, Gary Blasi, *Advocacy Against the Stereotype: Lessons from Cognitive Social Psychology*, 49 UCLA L. REV. 1241, 1250 (2002); Josie Foehrenbach Brown, *Escaping the Circle by Confronting Classroom Stereotyping: A Step toward Equality in the Daily Educational Experience of Children of Color*, 11 ASIAN L.J. 216, 228 n.65 (2004); Lateef Mtima, *The Road to the Bench: Not Even Good (subliminal) Intentions*, 8 U. CHI. L. SCH. ROUNDTABLE 135, 156 (2001); Pollard, *supra* note 2, at 918, 918 n.18; Reshma M. Saujani, "The Implicit Association Test": A Measure of Unconscious Racism in Legislative Decision-Making, 8 MICH. J. RACE & L. 395, 413–15 (2003); Michael S. Shin, *Redressing Wounds: Finding a Legal Framework to Remedy Racial Disparities in Medical Care*, 90 CALIF. L. REV. 2047, 2066 (2002).

¹⁵ Because other scholarship has thoroughly reviewed much of the broader work on unconscious and implicit bias, we focus only on scholarship related to the IAT. For reviews, see Kang, *supra* note 1; Levinson, *Forgotten Racial Equality*, *supra* note 1; 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) [hereinafter *Different Shades of Bias*].

¹⁶ As psychologists Nilanjana Dasgupta and Anthony Greenwald summarize, "[w]hen 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." 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, 803 (2001).

The following is a detailed description of the way the IAT is typically conducted: Study participants, working on computers, press two pre-designated keyboard keys as quickly as possible after seeing certain words or images on the computer monitors. The words and images that participants see are grouped into meaningful categories. These categories require participants to “pair an attitude object (for example, Black or White . . .) with either an evaluative dimension (for example, good or bad) or an attribute dimension (for example, home or career, science or arts)”¹⁷ Participants complete multiple trials of the pairing tasks, such that researchers can measure how participants perform in matching each of the concepts with each other. For example, in one trial of the most well known IATs, participants pair the concepts Good-White together by pressing a designated response key and the concepts Bad-Black together with a different response key. After completion of the trial, participants then pair the opposite concepts with each other, here Good-Black and Bad-White.¹⁸ The computer software that gathers the data¹⁹ measures the number of milliseconds it takes for participants to respond to each task. Scientists can then analyze (by comparing reaction times and error rates using a statistic called “D-prime”²⁰) whether participants hold implicit associations between the attitude object and dimension tested. Results of IATs conducted on race consistently show that “white Americans express a strong ‘white preference’ on the IAT.”²¹

As a measure, the IAT is quite flexible. Researchers have created dozens of different kinds of IATs. Some examples include: Gender-Science IAT, Gay-Straight IAT, Obama-McCain IAT, and the Fat-Thin IAT, among many others.²² The Gender-Science IAT, for example, requires participants to group together Male and Female photos with Science and Liberal Arts words. It is worth noting the flexibility of the IAT to test either evaluative dimension words (such as grouping Male-Female with Good-Bad), or attribute dimension words (such as grouping Male-Female with Career-Family). The IAT we created, the Guilty/Not Guilty IAT, requires participants to group together photos of White and Black faces (attitude-object photos) and Guilty and Not Guilty words (attribute

¹⁷ Levinson, *Forgotten Racial Equality*, *supra* note 1, at 355 (citing Banaji, *supra* note 4 at 117, 123).

¹⁸ Because participants may naturally be quicker at responding with one of their hands, participants complete these tasks twice, once for each response key, to eliminate differences based on hand preference. The order of the IAT tasks is also usually randomized to reduce order effects.

¹⁹ In our empirical study, we used the software Inquisit, produced by Millisecond Software.

²⁰ See Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm*, 85 J. PERSONALITY & SOC. PSYCHOL. 197, 201 (2003) [hereinafter *Improved Scoring Algorithm*].

²¹ Rachlinski et al., *Trial Judges*, *supra* note 5, at 1199. See also Levinson, *The Complicitous Mind*, *supra* note 14, at 612 (citing Brian Nosek et al., *Pervasiveness and Correlates of Implicit Attitudes and Stereotypes*, 18 EUR. REV. SOC. PSYCHOL. 36 (2008)).

²² See Project Implicit Website, <https://implicit.harvard.edu/implicit> (last visited October 30, 2010).

dimension words). As we will discuss, our empirical study of the IAT tested both the Guilty/Not Guilty IAT and the Pleasant/Unpleasant IAT,²³ an evaluative dimension IAT similar to the Good-Bad IAT.

B. *The IAT as a Symbol of Implicit Racial Bias*

Legal commentators have often recognized that racial discrimination in America has evolved from intentional and overt to unintentional and covert.²⁴ Reflecting the change in the way racial bias is practiced and propagated, legal scholarship considering implicit bias has most frequently focused on the ways in which these covert biases manifest in society, such as in hiring and promotion decisions.²⁵ In addition to explaining how the IAT and other social cognition measures reveal implicit bias in society, this scholarship considers the ways in which the law might react to the changing nature of discrimination.²⁶ In this subsection, we present a brief review of this legal scholarship. This summary underscores the critical importance of seeking to understand implicit bias in the law and highlights the need for projects that investigate the IAT in legally relevant settings.

Several scholars have relied on the IAT in proposing ways that implicit bias is relevant in the legal setting. In a fascinating project that introduced many legal scholars to the IAT, Jerry Kang relied on the IAT and other social cognition studies to argue that a Federal Communications Commission policy favoring local news may actually serve to propagate implicit bias in society.²⁷ Kang conducted a detailed review of a variety of compelling social cognition projects, called the results of IATs and other studies “stunning,” and urged that researchers pursue a broad research agenda in investigating implicit bias.²⁸

Since Kang’s 2005 project, scholars have considered other ways that the IAT and other measures might reflect inequality in society or the legal system. For

²³ See Anthony G. Greenwald et al., *Measuring Individual Differences in Implicit Cognition: The Implicit Association Test*, 74 J. PERSONALITY & SOC. PSYCHOL. 1464 (1998).

²⁴ See Krieger, *supra* note 1. See also Samuel R. Bagenstos, *Trapped in the Feedback Loop: A Response to Professor Days*, 49 ST. LOUIS U. L.J. 1007, 1009 (2005) (“[T]here is an emerging consensus that implicit or unconscious bias is becoming a more significant contributor to continuing workplace inequalities.”); Emily M.S. Houh, *Toward Praxis*, 39 U.C. DAVIS L. REV. 905, 909 (2006); McGinley, *supra* note 2, at 418 (noting that “the nature of discrimination has changed”).

²⁵ Levinson & Young, *Different Shades of Bias*, *supra* note 15, at 312–15.

²⁶ For various perspectives on how the law might react to the changing nature of discrimination, see Bagenstos, *The Structural Turn*, *supra* note 2; Blasi, *supra* note 14, at 1246–54; Krieger, *supra* note 1; Linda Hamilton Krieger & Susan T. Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 CALIF. L. REV. 997 (2006); Pollard, *supra* note 2; Susan Sturm, *Second Generation Employment Discrimination: A Structural Approach*, 101 COLUM. L. REV. 458, 459, 522–67 (2001).

²⁷ Kang, *supra* note 1.

²⁸ Kang, *supra* note 1, at 1493, 1536–38.

example, in separate projects, Justin Levinson and Antony Page relied on the IAT and other social cognition studies in examining legal decision-making. Levinson critiqued the ways people misremember information, and argued that judges and jurors may misremember case facts in racially biased ways.²⁹ Page argued that attorneys might unintentionally rely on implicit biases when using peremptory challenges.³⁰

Several scholars have even engaged in interdisciplinary collaborations to increase the scientific sophistication of their analysis and to bolster their claims. Linda Krieger, who famously introduced pre-IAT social cognition research to employment discrimination scholars in the mid 1990s,³¹ teamed up with social psychologist Susan Fiske and IAT co-creator Anthony Greenwald in separate projects that further evaluated implicit social cognition in the law.³² Building on claims made by Krieger before IAT research became mainstream, Krieger and Fiske argued that employment discrimination law must change to account for the changing nature of discrimination.³³ Greenwald and Krieger explained the IAT in great detail, and presented evidence that the IAT serves as a meaningful predictor of behavior.³⁴ Focusing on the changing nature of the debate over affirmative action, Kang collaborated with another IAT co-creator, Mazharin Banaji.³⁵ Proposing a new model of affirmative action called “fair measures,” Kang and Banaji relied on the IAT and other social cognition research to argue that implicit biases reflect continuing societal inequality that must be remedied.³⁶

All of the studies discussed represent just some of a rapidly growing field; more and more researchers claim that the IAT is a valid indicator that implicit racial bias is present in society and the legal system.³⁷ Yet the progress in

²⁹ Levinson, *Forgotten Racial Equality*, *supra* note 1.

³⁰ Page, *supra* note 1, at 160.

³¹ See Krieger, *supra* note 1.

³² See Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CALIF. L. REV. 945 (2006); Krieger & Fiske, *supra* note 26.

³³ Krieger & Fiske, *supra* note 26, at 1027–61. Krieger and Fiske suggested that the law “extract from normative legal reasoning the intuitive social science already there and to subject it to empirical scrutiny.” *Id.* at 1061.

³⁴ Greenwald & Krieger, *supra* note 32, at 952–55.

³⁵ See Jerry Kang & Mahzarin R. Banaji, *Fair Measures: A Behavioral Realist Revision of “Affirmative Action,”* 94 CALIF. L. REV. 1063 (2006).

³⁶ See *id.* at 1090–1110.

³⁷ Debra Lyn Bassett, *The Rural Venue*, 57 ALA. L. REV. 941, 944–45 (2006) (discussing the rural venue and stereotypes); Gary Blasi & John T. Jost, *System Justification Theory and Research: Implications for Law, Legal Advocacy, and Social Justice*, 94 CALIF. L. REV. 1119, 1137 (2006) (considering the IAT and implicit biases in the context of system justification theory); Ivan E. Bodensteiner, *The Implications of Psychological Research Related to Unconscious Discrimination and Implicit Bias in Proving Intentional Discrimination*, 73 MO. L. REV. 83 (2008) (considering standards of proof in light of social cognition work); Paul Butler, *Rehnquist, Racism, and Race Jurisprudence*, 74 GEO. WASH. L. REV. 1019, 1035–1037, 1041 (2006) (suggesting that former United States Chief Justice William Rehnquist may have harbored implicit racial biases, and further

incorporating the IAT into legal scholarship has not been limitless. Scholars have still yet to connect specifically the IAT measure with legal claims and have generally been satisfied to stick to fairly broad statements regarding how implicit bias manifests in the legal setting. Commentators have thus left open the question of whether the IAT itself can demonstrate anything compelling enough to legitimize legal changes. Most have argued that the compelling evidence of hundreds of rigorously tested implicit social cognition studies, including but not limited to the IAT, legitimizes some forms of legal or social change.³⁸ Yet others have been hesitant to accept the idea that legal change should be predicated on a measure of something that is largely outside of a person's conscious control.³⁹ And the debate seems to be heating up still.⁴⁰ Notwithstanding this debate, a large number of projects discussing the IAT and implicit bias continue to emerge in legal scholarship.⁴¹

suggesting that the IAT could serve as a meaningful test in evaluating judicial nominees); David L. Faigman et al., *A Matter of Fit: The Law of Discrimination and the Science of Implicit Bias*, 59 HASTINGS L.J. 1389 (2008) (focusing on the role of implicit bias in the workplace, especially as it is related to expert testimony on implicit bias); Christine Jolls & Cass R. Sunstein, *The Law of Implicit Bias*, 94 CALIF. L. REV. 969 (2006); Levinson, *The Complicitous Mind*, *supra* note 14; Justin D. Levinson, *Culture, Cognitions, and Legal Decision-Making*, in HANDBOOK OF MOTIVATION AND COGNITION ACROSS CULTURES 423–39 (Richard M. Sorrentino & Susumu Yamaguchi eds., 2008); Russell K. Robinson, *Perceptual Segregation*, 108 COLUM. L. REV. 1093 (2008) (proposing a theory of perceptual segregation). *See also* Dale Larson, Comment, *Unconsciously Regarded as Disabled: Implicit Bias and the Regarded-As Prong of the Americans With Disabilities Act*, 56 UCLA L. REV. 451 (2008).

³⁸ *See, e.g.*, Kang, *supra* note 1; Kang & Banaji, *supra* note 35; Krieger & Fiske, *supra* note 26; Levinson, *Forgotten Racial Equality*, *supra* note 1; Page, *supra* note 1.

³⁹ Mitchell & Tetlock, *Perils of Mindreading*, *supra* note 8; Amy L. Wax, *The Discriminating Mind: Define It, Prove It*, 40 CONN. L. REV. 979 (2008).

⁴⁰ *See, e.g.*, Bagenstos, “Science” and Antidiscrimination Law, *supra* note 8; Adam Benforado & Jon Hanson, *Legal Academic Backlash: The Response of Legal Theorists to Situationist Insights*, 57 EMORY L.J. 1087, 1135–43 (2008); Kristin A. Lane et al., *Implicit Social Cognition and Law*, 3 ANN. REV. L. & SOC. SCI. 427 (2007); Mitchell & Tetlock, *Facts Do Matter*, *supra* note 8.

⁴¹ *See, e.g.*, Richard Delgado & Jean Stefancic, *Four Observations About Hate Speech*, 44 WAKE FOREST L. REV. 353, 365–66 (2009) (suggesting that hate speech may lead to implicit bias); Alex Geisinger, *Rethinking Profiling: A Cognitive Model of Bias and Its Legal Implications*, 86 OR. L. REV. 657, 658 (2007) (claiming that racial profiling relies on cognitive processes that harbor implicit biases); Tristin K. Green & Alexandra Kalev, *Discrimination-Reducing Measures at the Relational Level*, 59 HASTINGS L.J. 1435 (2008) (considering the relational aspects of implicit bias in the workplace); Jonathan Kahn, *Race, Genes, and Justice: A Call to Reform the Presentation of Forensic DNA Evidence in Criminal Trials*, 74 BROOK. L. REV. 325, 373 (2008) (discussing the results of IATs and stating, “To the extent that such implicit race bias might already be present among average jurors, injecting race into the presentation of forensic DNA evidence presents a significant danger of tainting the proceedings with unfair prejudice.”); Cynthia Lee, *The Gay Panic Defense*, 42 U.C. DAVIS L. REV. 471, 479 (2008) (discussing implicit bias in the context of sexual orientation bias); Avital Mentovich & John T. Jost, *The Ideological “Id”? System Justification and the Unconscious Perpetuation of Inequality*, 40 CONN. L. REV. 1095 (2008) (considering Lawrence’s 1987 article, *supra* note 14, in an updated scientific perspective); Michael B. Mushlin & Naomi Roslyn Galtz, *Getting Real About Race and Prisoner Rights*, 36 FORDHAM URB. L.J. 27, 42–46 (2009)

The reliance on implicit social cognition research, and the IAT in particular, not only underscores major progress in legal scholarship, but also highlights a gap between implicit social cognition research and legal scholarship. Legal researchers often rely on the IAT for the proposition that people are implicitly biased, and tend to link it to a variety of legal claims. Yet few scholars have examined how the IAT may be used as a measure to test something specific in the legal setting. The next subsection reviews the small number of projects that have empirically examined the IAT in the legal context. These empirical studies highlight the potential of implementing tests such as the IAT in legally relevant contexts.

C. Law-Based Empirical Study of the IAT

Despite the growing familiarity with the IAT in legal scholarship, few research teams have empirically tested the IAT in a legal setting.⁴² In an early empirical study, Theodore Eisenberg and Sheri Lynn Johnson employed the IAT in a legal setting by testing whether capital defense attorneys displayed implicit racial bias.⁴³ Using a paper and pencil version of the Good/Bad Black/White IAT, the researchers found that the defense attorney participants, whom they expected to resist bias if at all possible, harbored strong implicit bias against African Americans.⁴⁴ Like other early tests of the IAT, however, the researchers did not test whether these implicit biases predicted the attorneys' decision-making. Nonetheless, their work demonstrated compelling results among a particularly noteworthy participant population.

In one of the only other law-focused empirical studies employing the IAT, Jeffrey Rachlinski collaborated with Johnson and others to test the ability of the

(discussing implicit bias in the context of prisoners' rights); Rigel C. Oliveri, *Between a Rock and a Hard Place: Landlords, Latinos, Anti-Illegal Immigrant Ordinances, and Housing Discrimination*, 62 VAND. L. REV. 55, 74–77 (2009) (predicting that implicit bias leads to housing discrimination against illegal immigrants); Gregory S. Parks & Quinetta M. Roberson, *Michelle Obama: A Contemporary Analysis of Race and Gender Discrimination through the Lens of Title VII*, 20 HASTINGS WOMEN'S L.J. 3 (2009) (considering race- and gender-based implicit bias in politics and in the workplace); Gregory S. Parks & Shayne E. Jones, "Nigger": A Critical Race Realist Analysis of the N-Word Within Hate Crimes Law, 98 J. CRIM. L. & CRIMINOLOGY 1305 (2008); Robert G. Schwemm, *Why Do Landlords Still Discriminate (And What Can Be Done About It)?*, 40 J. MARSHALL L. REV. 455, 507 (2007) (noting that implicit bias may affect housing rentals more than employment decisions).

⁴² Rachlinski et al., *Trial Judges*, *supra* note 5, at 1204. Rachlinski and his colleagues reported two studies in which researchers tested the predictive validity of the IAT in the legal setting. *Id.* (citing 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); 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)).

⁴³ See Theodore Eisenberg & Sheri Lynn Johnson, *Implicit Racial Attitudes of Death Penalty Lawyers*, 53 DEPAUL L. REV. 1539, 1542 (2004).

⁴⁴ *Id.* at 1545–48.

Good/Bad IAT to predict judicial decision-making.⁴⁵ Rachlinski and his colleagues recruited participants at judicial conferences, thus securing the unique opportunity to test a participant pool of willing judges.⁴⁶ The judge participants took a Good/Bad Black/White IAT, and then completed a task that asked them to make decisions in three hypothetical court scenarios.⁴⁷ In two of these scenarios, the race of the legal actor was ambiguous and was primed through a subliminal procedure.⁴⁸ In the third description, the race of the legal actor, White or Black, was varied.⁴⁹ The results of the study showed that, as in studies of the IAT in other populations, the judge participants displayed an implicit preference for White over Black.⁵⁰ That is, participants were faster to group together photos of White faces with Good words compared to Black faces with Good words. Next, the researchers found that the IAT results predicted responses in some, but not all of the judgment tasks.⁵¹ Specifically, they found that IAT results predicted racial bias

⁴⁵ See Rachlinski et al., *Trial Judges*, *supra* note 5. It is important to note the specific IAT that they employed, as different IATs may have different predictive abilities in various settings. For example, Laurie Rudman and Richard Ashmore tested whether the Good/Bad IAT and Stereotype IAT predicted participants' economic decisions. See Laurie A. Rudman & Richard D. Ashmore, *Discrimination and the Implicit Association Test*, 10 GROUP PROCESSES & INTERGROUP REL. 359 (2007). They found that the stereotype IAT, but not the Good/Bad IAT, predicted biases in economic decision-making. *Id.* at 368.

⁴⁶ The same researchers had previously used this technique to run empirical studies on judges. See Rachlinski et al., *Trial Judges*, *supra* note 5, at 1205 (citing Chris Guthrie et al., *Blinking on the Bench: How Judges Decide Cases*, 93 CORNELL L. REV. 1, 13 (2007); Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 814–15 (2001); 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)).

⁴⁷ Rachlinski et al., *Trial Judges*, *supra* note 5, at 1207–08.

⁴⁸ *Id.* at 1212. Rachlinski and his colleagues used a procedure similar to one conducted by Sandra Graham and Brian Lowery. *Id.* (citing Sandra Graham & Brian S. Lowery, *Priming Unconscious Racial Stereotypes About Adolescent Offenders*, 28 LAW & HUM. BEHAV. 483, 487–88 (2004)). Graham and Lowery's methodology was adapted from prior research conducted by Patricia Devine (Patricia G. Devine, *Stereotypes and Prejudice: Their Automatic and Controlled Components*, 56 J. PERSONALITY & SOC. PSYCHOL. 5, 7, 9, 13 (1989)), as well as John Bargh and Paula Pietromonaco (John A. Bargh & Paula Pietromonaco, *Automatic Information Processing and Social Perception: The Influence of Trait Information Presented Outside of Conscious Awareness on Impression Formation*, 43 J. PERSONALITY & SOC. PSYCHOL. 437, 439, 445 (1982)). The priming task used by Rachlinski and his colleagues did not specifically prime Black for one group and White for another, but Black for one group and ambiguous for the other. Rachlinski et al., *Trial Judges*, *supra* note 5, at 1213. It is also worth noting that the prime did not refer specifically to anything about the legal stories they used, and was designed more generally to prime all related knowledge structures and stereotypes of African Americans. Thus, it is unclear whether the results should be treated similarly to a situation where the race of the legal actors had been primed.

⁴⁹ Rachlinski et al., *Trial Judges*, *supra* note 5, at 1208.

⁵⁰ *Id.* at 1209–11.

⁵¹ *Id.* at 1211–19.

in some cases where the defendant's race had been subliminally primed, but not in cases where the race of the defendant was explicitly identified.⁵²

Rachlinski and his colleagues summarized their results as indicating that “judges, like the rest of us, possess implicit biases[,]”⁵³ but noted that “the judges managed, for the most part, to avoid the influence of unconscious biases when they were told of the defendant's race.”⁵⁴ They concluded, that among other things, “[t]he 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.”⁵⁵

As the reviewed research demonstrates, when examining both non-empirical and empirical applications of the IAT in the legal setting, two themes emerge: first, legal scholars have mostly confined themselves to discussing how existing IATs may or may not affect societal and legal decision-making; and second, the IAT has rarely been evaluated to see if it predicts legal decision-making. We addressed these themes in our empirical study, which tested a new IAT directly within the legal setting and examined whether the IAT results predicted decision-making. The next section explains why we created the Guilty/Not Guilty IAT and considers its importance in light of legal doctrine on the presumption of innocence.

III. DEVELOPING AN IAT FOR THE LEGAL SETTING

Although empirical research on implicit bias has been prominent in legal scholarship, there are significant opportunities to empirically investigate implicit bias directly within meaningful legal domains. Some fundamental legal principles, such as the presumption of innocence in criminal law, are particularly ripe for empirical testing, first, because they harbor deep legal meaning, and second, because they are formulated in a way that makes them testable.

A. A Law-Specific Measure of Bias

Scholars should strive to develop empirical measures that test legal concepts as directly as possible. If one wants to test whether people hold implicit associations between race and criminal guilt, one need not speculate: it is entirely possible to examine just that. This is particularly the case when legal concepts (such as Guilty/Not Guilty) hold deep societal meanings, both implicit and explicit, that are different than the psychological concepts that are regularly tested. Phrased another way, existing psychological measures may not measure the concepts the law cares about most. Although the examination of non-legal concepts (such as

⁵² *Id.* at 1219.

⁵³ *Id.* at 1232.

⁵⁴ *Id.*

⁵⁵ *Id.*

implicit attitudes of good-bad)⁵⁶ can reveal deep societal inequalities and raise questions of racial justice in the law (as scholars have consistently recognized⁵⁷), it may not always be optimal to generalize social science results to the legal setting when one can test legal concepts themselves.⁵⁸

Before designing a study, one should critically examine what that study might demonstrate in light of existing scholarship. When considering whether to develop an IAT to test whether people hold implicit associations between Black and Guilty, one should first ask whether existing IATs might already answer this question. At first glance, IATs testing implicit attitudes about race, such as good and bad, pleasant and unpleasant, do not reveal jurors' or mock jurors' implicit associations relating to a defendant's criminal guilt. Yet studies of predictive validity might offer some value here. It is possible to examine whether implicit associations of good and bad, for example, might predict Guilty and Not Guilty decisions. This type of predictive validity study is desirable in that it relies on already validated and widely available IATs to assess predictive validity.⁵⁹ Such a study, including the one conducted by Rachlinski and his colleagues, involves evaluating whether there is a predictive relationship between implicit (good-bad) attitudes about race and jury decision-making.⁶⁰ If such a study convincingly demonstrated a link between implicit attitudes (Black-bad) and jury decision (Guilty), then one might conclude that implicit attitudes predict jury decisions in racially biased ways. Because of this benefit, in addition to creating and testing the Guilty/Not Guilty IAT, we also tested whether the Pleasant/Unpleasant IAT predicted mock juror evaluations of evidence.

Despite the benefits of predictive validity research, from a legal perspective there is a difference between testing predictive validity of implicit attitudes and testing implicit associations regarding specific legal assumptions such as the presumption of innocence. Although testing predictive validity can be eye-opening, a predictive validity study of an attitude-based IAT could not fully examine whether people hold implicit associations between Black and Guilty

⁵⁶ Attitude targets, such as "good" and "bad", may reflect important societal principles, but attitudes alone rarely have relevance in the legal system. Thus, to the extent that law holds dichotomous constructs with independent meaning, such as Guilty/Not Guilty, the IAT is one proper measure that can be used. Other similar measures include the Go/No-Go Association Task. See Brian A. Nosek & Mahzarin R. Banaji, *The Go/No-Go Association Task*, 19 SOC. COGNITION 625 (2001). This task, conceptually similar to the IAT, allows for researchers to test individual target groups with dichotomous pairs. *Id.* at 627.

⁵⁷ See *supra* Section II.B and accompanying text.

⁵⁸ It is not always easy to directly test legal concepts using the IAT. Because the IAT was specifically designed to test dichotomous principles, the legal concept of Guilty/Not Guilty fits particularly well into the structure of the test. Other dichotomous legal concepts could similarly be tested using IAT.

⁵⁹ See, e.g., Rachlinski et al., *Trial Judges*, *supra* note 5; Rudman & Ashmore, *supra* note 45.

⁶⁰ Rachlinski and colleagues have published the only study doing this. See Rachlinski et al., *Trial Judges*, *supra* note 5. Rachlinski and colleagues cite an unpublished study by Livingston, *supra* note 42, that they describe as linking the IAT to mock jury decisions. *Id.* at 1204.

compared to White and Guilty. Examining these specific implicit associations is particularly compelling in light of the presumption of innocence.

B. *The (Implicit) Presumption of Guilt*

The most compelling reason for testing implicit associations of guilt is that the law espouses a presumption of innocence, a presumption that may be contradicted by racial cues and the realities of the human mind. The presumption of innocence is a fundamental principle of criminal law.⁶¹ According to the United States Supreme Court in *Coffin v. United States*, “[t]he principle that there is a presumption of innocence in favor of the accused is the undoubted law, axiomatic and elementary, and its enforcement lies at the foundation of the administration of our criminal law.”⁶²

In light of the vast importance of the presumption of innocence, partially underscored by the widespread public understanding that a defendant is “innocent until proven guilty,” it is not surprising that scholars have long questioned whether there is a relationship between racial inequality and the presumption of innocence. In 1883, Frederick Douglas surmised, “The reasonable doubt which is usually interposed to save the life and liberty of a white man charged with crime, seldom has any force or effect when a colored man is accused of crime.”⁶³ More contemporary discussions of racial bias in criminal law underscore the potentially deep connections between racial bias and the presumption of innocence.⁶⁴ Although none of these projects have empirically tested implicit associations of race and criminal guilt, some scholars have considered that racial disparities in conviction and sentencing may be driven by racial bias.⁶⁵ Robert Entman and

⁶¹ Scholars have traced the presumption of innocence back to ancient Hebrew law, Roman law, and medieval twelfth-century Italian law. See Rinat Kitai, *Presuming Innocence*, 55 OKLA. L. REV. 257, 260–62 (2002). See also Kenneth Pennington, *Innocent Until Proven Guilty: The Origins of a Legal Maxim*, 63 JURIST 106 (2003) (providing a detailed account of the origins of the presumption of innocence).

⁶² *Coffin v. United States*, 156 U.S. 432, 453 (1895). See also *Estelle v. Williams*, 425 U.S. 501, 503 (1976) (“The presumption of innocence . . . is a basic component of a fair trial under our system of criminal justice.”). According to Scott Sundby, “the presumption of innocence is given vitality primarily through the requirement that the government prove the defendant’s guilt beyond a reasonable doubt.” Scott E. Sundby, *The Reasonable Doubt Rule and the Meaning of Innocence*, 40 HASTINGS L.J. 457, 458 (1989).

⁶³ Sheri Lynn Johnson, *Black Innocence and the White Jury*, 83 MICH. L. REV. 1611, 1612–13 (1985) (quoting Frederick Douglass as quoted in *Symposium on The Black Lawyer in America Today*, HARV. L. SCH. BULL., Feb. 1971, at 6, 57).

⁶⁴ See *id.*; KENNEDY, *supra* note 11. See also Montré D. Carodine, “*The Mis-Characterization of the Negro*”: *A Race Critique of the Prior Conviction Impeachment Rule*, 84 IND. L.J. 521, 527 (2009) (“When Blacks are unfairly ‘taxed’ in the criminal system with perceived criminality, Whites receive an undeserved ‘credit’ with a perceived innocence or worthiness of redemption.”).

⁶⁵ See Bryan K. Fair, *Using Parrots to Kill Mockingbirds: Yet Another Racial Prosecution and Wrongful Conviction in Maycomb*, 45 ALA. L. REV. 403, 408 (1994) (“It is misguided to believe

Kimberly Gross, for example, posited that “black defendants in criminal cases are especially likely to be presumed guilty because they are subject to the stereotypes or heuristics that most whites apply to the category ‘black person.’”⁶⁶ These scholars all maintain a shared perspective underlying our hypothesis: In a criminal justice system that reveals continuing massive racial disparities,⁶⁷ it is quite possible that African American defendants may be afforded a weaker presumption of innocence than White defendants. Driven by this research question, we set out to test it empirically.

IV. THE GUILTY/NOT GUILTY IAT: AN EMPIRICAL STUDY

This section details the empirical study we conducted. It first explains the study we designed and then presents the results of the study.

A. *Methods and Materials*

Participants in the empirical study were sixty-seven jury eligible undergraduate and graduate students at the University of Hawaii who participated in the study for extra course credit.⁶⁸ After giving informed consent, participants

that White folks can discard strongly held negative attitudes about Blacks when Whites act as police, jurors, lawyers, or judges in criminal cases with a Black criminal defendant. . . . Once we admit racial animus into the courtroom, we abandon the presumption of innocence standard that is supposedly central to our jurisprudential traditions.”). Within interdisciplinary scholarship, several projects have investigated whether jurors hold racial biases using mock jury studies. Some scholars have aggregated these studies in order to conduct large scale quantitative or qualitative reviews. See Johnson, *supra* note 63; Nancy J. King, *Postconviction Review of Jury Discrimination: Measuring the Effects of Juror Race on Jury Decisions*, 92 MICH. L. REV. 63 (1993); Tara L. Mitchell et al., *Racial Bias in Mock Juror Decision-Making: A Meta-Analytic Review of Defendant Treatment*, 29 LAW & HUM. BEHAV. 621 (2005); Samuel R. Sommers & Phoebe C. Ellsworth, *How Much Do We Really Know About Race and Juries? A Review of Social Science Theory and Research*, 78 CHI-KENT L. REV. 997 (2003).

⁶⁶ Robert M. Entman & Kimberly A. Gross, *Race to Judgment: Stereotyping Media and Criminal Defendants*, 71 LAW & CONTEMP. PROBS. 93, 97 (2008). Several other articles at least reference the idea that the presumption of innocence may be racially biased. See, e.g., Carodine, *supra* note 64; Fair, *supra* note 65; Robin K. Magee, *The Myth of the Good Cop and the Inadequacy of Fourth Amendment Remedies for Black Men: Contrasting Presumptions of Innocence and Guilt*, 23 CAP. U. L. REV. 151, 153–55 (1994).

⁶⁷ The U.S. Department of Justice Bureau of Justice Statistics reports that more than 6 in 10 inmates in local jails are minorities, including 41% “black.” U.S. DEP’T OF JUSTICE BUREAU OF JUSTICE STATISTICS, PROFILE OF JAIL INMATES 1996 1, 3 (1998), available at <http://bjs.ojp.usdoj.gov/content/pub/pdf/pji96.pdf>. In addition, the Bureau reports that the lifetime likelihood of going to state or federal prison is 18.6% for “blacks” compared to 10% for Hispanics and just 3.4% for “whites.” U.S. DEP’T OF JUSTICE BUREAU OF JUSTICE STATISTICS, PREVALENCE OF IMPRISONMENT IN THE U.S. POPULATION, 1971–2001 1, 8 (2003), available at <http://bjs.ojp.usdoj.gov/content/pub/pdf/pius01.pdf>.

⁶⁸ Data from one participant was excluded because it was incomplete. All but seven of the participants, who were residents of other states, were jury eligible in Hawai‘i. All participants were

began the all-computerized task in a laboratory with two separate cubicles, each containing a Dell desktop computer. Participants completed several measures, including: a (Black/White) Guilty/Not Guilty IAT that we developed, a (Black/White) Pleasant/Unpleasant IAT,⁶⁹ the Modern Racism Scale, feeling thermometers, and a robbery evidence evaluation task.⁷⁰ Participants completed the robbery evidence evaluation task first, and then completed the remaining tasks, including the IATs, in randomized order. Participants provided demographic information at the end of the study.

Participants completed the two IATs in counterbalanced order.⁷¹ One IAT was the Guilty/Not Guilty IAT measure we developed: a race IAT with the attribute concepts of Guilty and Not Guilty, and target concepts of Black and White.⁷² The other IAT, the Pleasant/Unpleasant IAT, was a race IAT with the evaluative concepts of Pleasant and Unpleasant, and target concepts of Black and White.⁷³ The Modern Racism Scale consists of a series of questions that measure

over 18 years of age and none had been convicted of a felony. Participant ages ranged from 18 to 40, with a mean age of 21.85 (SD=3.95). Twenty-five participants identified themselves as Japanese American, eighteen participants identified themselves as European American, and five participants identified themselves as Chinese American. Other participants identified themselves as Native Hawaiian, Pacific Islander, Korean American, and Latino. Three participants identified themselves as “mixed race,” and five indicated “other.” The likely ethnic diversity of a Hawai‘i-based sample, as we found here, is notable. *See* Levinson, *Forgotten Racial Equality*, *supra* note 1, at 396 (noting the uniqueness and diversity of that Hawai‘i based sample, and considering the meaning of the study results in light of the diverse sample).

⁶⁹ The instructions for the IATs were as follows: “For the next set of tasks you will be shown words one at a time in the middle of the computer screen and asked to sort them into categories. Your task is to sort each item into its correct category as fast as you can by pressing EITHER the 'D' key or the 'K' key. IMPORTANT: Press the 'D' key using your left index finger, or 'K' key using your right index finger. The categories associated with the 'D' and 'K' keys will be shown at the top of each screen. Please pay close attention to these category labels—they change for each sorting task!”

⁷⁰ For more on this task, see Levinson & Young, *Different Shades of Bias*, *supra* note 15.

⁷¹ Also, within each IAT, the tasks were counterbalanced.

⁷² One of the particularly interesting elements of the IAT is the way in which target words are selected for inclusion into the dichotomous categories used in the IAT. In designing the Guilty/Not Guilty IAT, we needed to choose stimuli for Black, White, Guilty, and Not Guilty. Our first goal was to use stimuli that were validated, so as not to reinvent the wheel. For the categories of Black and White, we thus used stimuli used in other IATs. These stimuli consisted of six photographs of faces of White men and women and six photographs of faces of Black men and women. To select target words for the categories Guilty and Not Guilty, we conducted a pre-test to ensure that the terms we selected were most representative of the target concepts. This pre-test asked participants to rank how much various words were associated with the criminal law concepts of Guilty and Not Guilty. Considering mean score and standard deviation, we selected the target words that participants agreed upon most. For Guilty, the target words were: at fault, caught in the act, committed crime, convict, criminal, did it, perpetrator, responsible for crime. For Not Guilty, the target words were: acquitted, blameless, cleared of charges, didn't do it, did not commit crime, wrongfully accused, guilt free, and innocent.

⁷³ This is one of the more common IATs that measures implicit attitudes. *See* Greenwald et al., *Measuring Individual Differences*, *supra* note 4, at 1465. For Pleasant, the target words were

self reports regarding racial beliefs. It is one of the favored social cognition measures for evaluating explicit racial preferences.⁷⁴ The Scale asks participants to respond to several statements about Blacks in America.⁷⁵ The feeling thermometer measures, like the Modern Racism Scale, were designed to evaluate explicit racial preferences. The feeling thermometer measures asked: “How warm do you feel towards European Americans,” and “How warm do you feel towards African Americans?” The evidence evaluation task presented participants with the story of an armed robbery. After reading the story, participants then saw a series of crime scene photos, and were primed with either a photo of a dark skinned perpetrator or light skinned perpetrator.⁷⁶ Participants were then presented with a list of individual pieces of evidence, and asked to score each piece of evidence based on whether it tended to indicate that the defendant was guilty or not guilty.⁷⁷ Finally, participants were also asked to decide whether the defendant was guilty or not guilty, both on a dichotomous scale and on a continuous scale.⁷⁸ After completing the study, participants were thanked for their participation and debriefed.

B. Scoring the IAT

Because the IAT is such a complicated measure, social scientists have considered several scoring algorithms regarding the IAT. In calculating the results of our study, we relied on the updated scoring algorithms suggested by Greenwald and his colleagues.⁷⁹ These improved algorithms addressed several challenges that were raised regarding the original IAT scoring algorithm.⁸⁰

beautiful, lovable, valuable, attractive, and smart. For Unpleasant, the target words were ugly, useless, stupid, hostile, and inferior.

⁷⁴ See generally John B. McConahay, *Modern Racism, Ambivalence, and the Modern Racism Scale*, in PREJUDICE, DISCRIMINATION, AND RACISM 91 (John F. Dovidio & Samuel L. Gaertner eds., 1986) (introducing the Modern Racism Scale).

⁷⁵ For example, one statement that participants respond to is: “Discrimination against Blacks is no longer a problem in the United States.”

⁷⁶ The photos were identical, with the skin-tone altered using computer software. This skin-tone manipulation was a between-subjects independent variable. Half the participants saw a dark-skinned perpetrator and half the participants saw a light skinned perpetrator. Participants were randomly assigned to the two priming conditions. For more on this element of the study, including a theoretical discussion of evidence evaluation and detailed results, see generally Levinson & Young, *Different Shades of Bias*, *supra* note 15.

⁷⁷ Some examples of this evidence were: “the defendant was a youth Golden Gloves boxing champ in 2006; the defendant purchased an untraceable handgun three weeks before the robbery; the defendant is a member of an anti-violence organization; and the defendant had a used movie ticket stub for a show that started 20 minutes before the crime occurred.”

⁷⁸ The continuous scale ranged from 0 (definitely not guilty) to 100 (definitely guilty).

⁷⁹ See Greenwald et al., *Improved Scoring Algorithm*, *supra* note 20, at 213–15.

⁸⁰ Greenwald, Nosek and Banaji’s suggested improved scoring measure for the IAT, called a *D* score, has improved test-response detection (for instance, it throws out indiscriminate responses or

C. Results

1. Guilty/Not Guilty IAT

The results of the Guilty/Not Guilty IAT confirmed our hypothesis that there is an implicit racial bias in the presumption of innocence. Participants displayed a significant association between Black and Guilty⁸¹ compared to White and Guilty,⁸² producing a significant IAT effect.⁸³ These results suggest that participants held an implicit association between Black and Guilty.

2. Pleasant/Unpleasant IAT

The results of this study reproduced the results found in many attitude-race IATs, including Pleasant/Unpleasant IATs conducted by other researchers.⁸⁴ Participants displayed a significant association between Black-Unpleasant compared to White-Unpleasant. Specifically, participants associated Black and unpleasant words⁸⁵ significantly faster than Black and pleasant words,⁸⁶ resulting in a significant IAT effect.⁸⁷

3. Relationship of Implicit Measures to Explicit Measures

We next computed correlation coefficients in order to assess the relationship between the implicit measures (the two IATs) and the explicit measures (the Modern Racism Scale and the feeling thermometers). Interestingly, some implicit scores were correlated with explicit scores. IAT scores on the Pleasant/Unpleasant IAT were correlated with scores on the Modern Racism Scale,⁸⁸ such that

responses that indicate a lack of attention) and incorporates an inclusive standard deviation for all congruent trials (for instance, both the practice and test block of white-guilty and black-not guilty). *Id.* at 213. Mean latencies are computed for each block, and complimentary blocks are subtracted from each other (e.g., practice white-not guilty and black-guilty would be subtracted from practice white-guilty and black-not guilty). *Id.* at 214. These two difference scores are divided by their inclusive standard deviation score, and the average of these two scores is called *D*. *Id.* For a detailed summary of Greenwald and his colleagues' scoring algorithm, see Rachlinski et al., *Trial Judges*, *supra* note 5, at 1245–46.

⁸¹ Mean = 727.63ms. The IAT measures target results together, such that this result measures an implicit association between the pairs of Black-Guilty and White-Not Guilty, compared to the pairs of White-Guilty and Black-Not Guilty.

⁸² Mean = 800.16ms, $t = 2.68$, $p < 0.01$.

⁸³ (*D*) of 0.18 ($t = 3.36$, $df = 65$, $p < 0.01$).

⁸⁴ For a broad review of IAT results, see Nosek et al., *supra* note 6, at 105–11.

⁸⁵ Mean = 613.92ms.

⁸⁶ Mean = 690.93ms, $t = 3.29$, $p < 0.01$.

⁸⁷ (*D*) of 0.21 ($t = 4.11$, $df = 65$, $p < 0.01$).

⁸⁸ $r = 0.33$, $p < 0.01$.

participants who displayed greater implicit bias were also likely to report less favorable explicit attitudes towards African Americans. Correlations between the Pleasant/Unpleasant IAT scores and scores on the feeling thermometer towards African Americans were marginally significant,⁸⁹ such that participants who displayed greater implicit bias were more likely to report cooler attitudes towards African Americans.

4. Warm Feelings Towards African Americans Correlated with Guilty Implicit Bias

Interestingly, correlation coefficients on the Guilty/Not Guilty IAT showed one significant correlation that differed from the correlations found on the Pleasant/Unpleasant IAT. Scores on the Guilty/Not Guilty IAT correlated with scores on the feeling thermometer regarding feelings towards African Americans,⁹⁰ such that people who reported feeling warmly towards African Americans were more likely to show an implicit guilty bias against Blacks.⁹¹

The opposite correlation patterns of the Guilty/Not Guilty IAT and the Pleasant/Unpleasant IAT suggest that the two implicit measures are measuring different phenomena and tap into different implicit constructs. Although the Pleasant/Unpleasant IAT appeared to tap into a construct measured by the Modern Racism Scale and (marginally) the feeling thermometer, the Guilty/Not Guilty IAT correlated with the feeling thermometer in an opposite direction and did not correlate with the Modern Racism Scale.

Table 1: Means, Standard Deviations of explicit and implicit measures, and correlations among them (n=66).

Variables	Means	SD	Correlations		
			1	2	3
1. Pleasant/Unpleasant IAT	0.21*	0.42			
2. Guilty/Not Guilty IAT	0.18*	0.42	-0.07		
3. Modern Racism Scale	12.56	4	0.33*	-0.01	
4. Thermometer_Black	6.18	1.48	-0.23+	0.29*	0.22+

Note: * $p < .05$ + $p < .10$.

⁸⁹ $r = -0.23, p = 0.06$.

⁹⁰ $r = 0.29, p < 0.05$.

⁹¹ Guilty/Not Guilty IAT scores were not significantly correlated with scores on the Modern Racism Scale, $r = -0.01, n = 66, p = 0.80$, indicating that using that particular measure does not relate to implicit biases in Guilty/Not Guilty associations.

5. Predictive Validity: Guilty/Not Guilty and Pleasant/Unpleasant IATs Predicted Evidence Judgments

To examine whether the IATs predicted participants' judgments of evidence and criminal guilt, we created a regression model investigating the impact of the two IATs on the totals of evidence judgments.⁹² The regression model showed that having stronger implicit associations between Black and Guilty,⁹³ and having higher Black/Unpleasant IAT scores,⁹⁴ predicted judgments of ambiguous evidence as more indicative of guilt.⁹⁵ It should be noted that the Guilty/Not Guilty and Pleasant/Unpleasant IATs were individually significant in this model, which indicates that each IAT separately predicted evidence judgments. In addition, because the two IATs were each significant, it also suggests that the Guilty/Not Guilty IAT and Pleasant/Unpleasant IAT are measuring different constructs and are not interchangeable as measures of implicit bias.

Table 2: Regression Analysis

	B	SE	<i>t</i>	<i>p</i>
<i>Constant</i>	88.58	4.11	21.54	.000
<i>BW</i>	5.54	2.49	2.22	.030
<i>GI</i>	6.61	2.96	2.23	.029
<i>AI</i>	9.11	3.01	3.03	.004

Note: BW= black vs. white (1=black; 0=white)

GI= guilty IAT

AI= attitude IAT

Regression equation: Evidence = 88.58 + 5.74 x BW + 6.61 x GI + 9.11 x AI + *e*

⁹² The regression model also included the variable of perpetrator skin tone (dark v. light) from the evidence evaluation task. This variable was also predictive of judgments of ambiguous evidence as more indicative of guilt, $\beta = 0.25$, $t = 2.22$, $p < 0.05$. This particular result is discussed in more detail by Levinson & Young, *Different Shades of Bias*, *supra* note 15, at 337.

⁹³ $\beta = 0.25$, $t = 2.23$, $p < 0.05$.

⁹⁴ $\beta = 0.34$, $t = 3.03$, $p < 0.01$.

⁹⁵ As expected, IAT score did not alone predict guilty verdicts or guilty scale judgments. In addition, it should be noted that our regression analysis found that the IATs predicted total evidence judgments, rather than total evidence judgments based upon the skin tone of the perpetrator. This later type of predictive validity should be measured in future studies with larger samples. Due to the small sample size in this study, we did not expect to find significant results on such a regression analysis.

In order to investigate the relationship between the two IATs further, we checked to see if the two IATs were correlated. The results indicated that the two IATs were not correlated with each other,⁹⁶ further suggesting that the Guilty/Not Guilty IAT measures a different construct than the Pleasant/Unpleasant IAT.

V. DISCUSSION OF RESULTS: IMPLICIT BIAS AND GUILTY AFRICAN AMERICAN MEN

The results of the empirical study show that, when it comes to racial equality and the presumption of innocence, there is reason for concern. First, we found that participants held implicit associations between Black and Guilty. Second, we found that these implicit associations were meaningful—they predicted judgments of the probative value of evidence. Third, we found that the Guilty/Not Guilty IAT was unrelated to and operated differently than a well established attitude-based IAT. Finally, we found that implicit attitudes of race and guilt are quite different from attitudes of race revealed by using explicit measures; in fact, one explicit measure even showed opposite results: participants who reported feeling warmer towards African Americans actually showed more bias on the Guilty/Not Guilty IAT.⁹⁷ These findings, taken together, raise questions about racial justice in the law and present evidence to challenge the integrity of the presumption of innocence. It goes without saying that this fundamental legal principle, which “lies at the foundation of the administration of our criminal law[,]”⁹⁸ should neither bend nor break when defendants are Black men.

It is also worth highlighting the success of the attitude IAT in our study. Like the Guilty/Not Guilty IAT, the Pleasant/Unpleasant IAT revealed a significant implicit racial bias and predicted evidence judgments. Although we believe that the Guilty/Not Guilty IAT holds the most promise in this particular legal domain, the versatility of attitude IATs is impressive. Researchers should continue to investigate a broad range of IATs in the legal setting.⁹⁹

Due to the single empirical study and still emerging nature of the Guilty/Not Guilty IAT, it is premature to discuss solutions, if there are any,¹⁰⁰ to the implicit

⁹⁶ $r = -0.07$, $p = 0.51$.

⁹⁷ Thus, asking jurors whether they can be unbiased is unlikely to reveal jurors with strong implicit biases.

⁹⁸ *Coffin v. United States*, 156 U.S. 432, 453 (1895).

⁹⁹ This future research should include stereotype IATs, which we did not test here. For one interesting study testing the predictive validity of attitude IATs and stereotype IATs, see Rudman & Ashmore, *supra* note 45, at 368 (finding that the stereotype IAT, but not the attitude IAT, predicted economic decisions based on group status).

¹⁰⁰ We intentionally phrase this qualifier in a pessimistic way. Although there is some reason to believe that implicit biases can be temporarily mitigated using certain techniques, none of these techniques address the cultural factors that lead to a society that harbors such biases. For more on debiasing, see Kang & Banaji, *supra* note 35, at 1107–08 (examining their proposal of “fair measures,” which relies heavily on encouraging counterstereotypic job holders); Levinson, *Forgotten*

bias revealed by the study. As a measure, the Guilty/Not Guilty IAT would be best served by follow-up testing and subsequent amelioration, if necessary. For example, the measure should be tested in a variety of locations and on broader populations. In addition, the predictive validity of the measure should be tested in realistic trial settings and in larger samples.

VI. CONCLUSION

We consider this study just one part of a broader effort to investigate whether implicit biases facilitate societal and legal inequality.¹⁰¹ It is our hope that future empirical studies of implicit bias in the law will continue this effort, specifically by investigating a broad range of legal domains where implicit bias may affect legal decision-making. There are so many areas of the law that have yet to be considered as possible hideouts for implicit bias. A broad range of legal areas, including but not limited to immigration law, contract law, and property law may unknowingly be functioning with the covert and dangerous help of implicit bias. We hope that future studies, particularly those conducted by interdisciplinary research teams, will pursue these areas, while also continuing to investigate the Guilty/Not Guilty IAT.

Racial Equality, *supra* note 1, at 411–412 (suggesting more diverse juries, multiculturalism training for juries, and a more counter-stereotypic community of lawyers, but noting that cultural change is the only true solution for implicit racial bias in the legal setting). For more on taking cultural responsibility for implicit bias, see generally Levinson, *The Complicitous Mind*, *supra* note 14.

¹⁰¹ This broader effort holds deep importance for those concerned with social justice. Our study, as well as many others, continues to raise questions about what implicit bias says about American culture and society. We do not believe that the covert and automatic nature of implicit racial biases, in light of persisting societal disparities, absolves society of the responsibility for their harm. See generally Charles Lawrence, *Unconscious Racism Revisited: Reflections on the Impact and Origins of “The Id, The Ego, and Equal Protection,”* 40 CONN. L. REV. 931 (2008); Levinson, *Forgotten Racial Equality*, *supra* note 1; Levinson, *The Complicitous Mind*, *supra* note 14.