

UCLA

UCLA Public Law & Legal Theory Series

Title

A Future History of Implicit Social Cognition and the Law

Permalink

<https://escholarship.org/uc/item/0h54x4rj>

Journal

UCLA School of Law Public Law & Legal Theory Series

Authors

Kang, Jerry

Lane, Kristin

Publication Date

2009

A Future History of Implicit Social Cognition and the Law

Jerry Kang* & Kristin Lane**

Version 3.1

8/12/2009 3:44 PM

© by authors

Please do not quote, cite, copy, or distribute further,
without explicit permission.

* Professor of Law, UCLA School of Law. <kang@law.ucla.edu> <<http://jerrykang.net>>. ** Assistant Professor of Psychology, Bard College. Thanks to: Ken Karst, Doug Kysar, Russell Robinson, Len Rubinowitz, Ed Stein, Michael Sullivan. Special thanks to Mahzarin Banaji.

Research assistance provided by: Kevin Minnick, Vivek Mittal, Jennie Park, Emily Reitz, Halpert Xiao, and Yi (Jenny) Xiao, Hentyle Yapp. Thanks to Hugh & Hazel Darling Law Library at UCLA School of Law.

Presented at: Emory Legal Theory Workshop; Yale Work Seminar; Southwestern University School of Law; CAPALF 2008 Keynote, Denver; University of California Berkeley Law School's Thelton Henderson Center Conference "Reclaiming & Reframing the Dialogue on Race and Racism;" Ohio State University School of Law; UCLA School of Law Summer Workshop Series; UCLA Critical Race Studies Faculty Seminars; Northwestern Critical Race Theory Colloquium.

Supported in part by: UCLA School of Law, UCLA Asian American Studies Center, UCLA Institute for American Cultures.

Abstract: The science of “implicit social cognition” (ISC) has demonstrated that we have implicit biases, in the form of stereotypes and attitudes that we are unaware of. Nevertheless, these implicit biases can alter our behavior, including how we might give an interview, hire a candidate, or even shoot a gun. As evidence mounts on this front, the law will have no choice but to respond to this revised understanding of human behavior and decisionmaking. To accelerate on-the-merits analysis, we offer a “future history” of how a new scientific consensus might be reached and how it might be incorporated into the law. First, this Article provides a current, concise, and accurate primer on the science. Second, it elucidates a call for “behavioral realism,” which asks that the law account for the most accurate model of human thought, decisionmaking, and action provided by the mind sciences. Third, it predicts critical markers in our future, which we label denial, minimization, politicization, resignation, and justification. This stylized future history provides us an analytic vocabulary for intellectual engagements incipient and soon-to-come; avoids unhelpful confusions, misreadings, or strawmen; answers reasonable objections to the extent that current scientific and legal understandings permit; and points us toward a broader research agenda for the next decade.

Introduction	1
I. Science Primer	3
A. The Psychological Context	3
B. The Challenge of Implicit Social Cognition	5
C. The Techniques	6
D. The Findings.....	8
1. Pervasiveness.....	9
2. Dissociation.....	11
3. Real-world Behaviors.....	12
II. Behavioral Realism	14
III. The Future History	16
A. Denial	16
1. Scientific Denial	16
2. Definitional Denial	20
B. Minimization	21
1. No real-world behaviors.....	21
2. Easy correction	25
C. Politicization.....	27
1. Overreaching by the Left	27
2. Backlash from the Right	29
D. Resignation.....	31
E. Justification	34
1. Descriptive	35
a. Accurate probability data.....	35
b. Rational processing	36
2. Normative	38
Conclusion.....	39

“Whatever may have been the extent of psychological knowledge at the time of *Plessy v. Ferguson*, this finding is amply supported by modern authority.”

—Brown v. Board of Education
347 U.S. 483, n.11 (1954)

“Doubt is our product, since it is the best means of competing with the ‘body of fact’ that exists in the mind of the general public. It is also the means of establishing a controversy.”

—Brown and Williamson
tobacco company memorandum (1969)

INTRODUCTION

[1] We start with three fictional stories.

[2] *Juan’s Interview.* Juan was desperate for the job. But he didn’t get it. Crushed, Juan grabs a beer with his friend. Juan grouches how hard he’s hustled all his life, but that somehow things always seem stacked against him. He says, “From the moment that interview started, it wasn’t gonna happen. That guy just didn’t like the look of me.” Juan is Mexican-American. His friend says “Stop whining.... there were lots of applicants. Just the luck of the draw. You’ll get the next one.” Juan shakes his head and says, “You just don’t get it.”

[3] *Skip’s Encounter.* After a long flight, Skip finally arrives home late at night, but his front door is jammed. With the help of his driver, they jimmy it open. Just as he lays down his bags, exhausted, he notices outside his door a police officer, who asks him to step outside. It’s his home, and Skip yells out that he’s a professor at the nearby campus, and that everything’s fine. The officer repeats, more sternly, to step outside, and moves his hand closer to his revolver. The officer expects deference. Skip is African American, and he is tired of deferring. He wonders why he’s being scrutinized. Disgusted, he reaches quickly for his wallet to offer ID, and yells “Take it!” The cop flinches and draws his weapon. Skip’s face is now pale, as he stares down the barrel of a gun. “He says slowly, it’s my faculty ID.....”¹

[4] *Grace’s Hire.* The faculty supporting Grace’s candidacy make their final pitch of the hiring meeting. They emphasize that they live in Southern California, but don’t have a single Asian American on the faculty. They say that Grace is eminently qualified. They say that she would also be a “role model” for minorities and women. Some add that she would also counter the stereotypes of Asian women: “She’s eloquent, charismatic, a firecracker! Would do us and our students good.” Those faculty opposed say Grace is good, but that Bill is

¹ This is fiction. Most important, when Professor Skip Gates was arrested at his home, the arresting officer did not draw his gun.

even better. More important, race should have nothing to do with academic merit, especially in a field like corporate law. It's wrong, unnecessary for Asians, and it's probably illegal. Bill happens to be White and liberal, like most of the faculty. He gets hired.

[5] How do you respond to these stories? No doubt your reactions turn on your political and social values. But they also depend on your understanding of the facts. What *really* happened in Juan's interview? Who's to blame in Skip's testy encounter? Would Grace's hire really "do us good," or is that just wishful thinking? In other words, even when we put aside philosophical and moral questions about equality, justice, and fairness, we still run into difficult empirical questions of how and whether race and gender matter.

[6] Thankfully, there has been a recent explosion of knowledge on the empirical front. At the nexus of social psychology, cognitive psychology, and cognitive neuroscience has emerged a new body of science called "implicit social cognition" (ISC). This field focuses on mental processes that affect social judgments but operate without conscious awareness or conscious control.² Nevertheless, these implicit thoughts and feelings leak into everyday behaviors such as whom we befriend, whose work we value, and whom we favor—withstanding our ignorance of any such influence. These discoveries, disseminated from hundreds of research programs, have provoked surprise, fascination, and even anger.

[7] As behavioral realists, we believe such findings—or in the parlance of *Brown v. Board of Education*, this "modern authority"—should have consequences. "Behavioral realism" insists that the law account for the most accurate model of human thought, decisionmaking, and action provided by the sciences. Theories and data from implicit social cognition are sharpening that model, and as a new scientific consensus emerges, the law should respond. Either the law should change to reflect that more accurate model, or provide reasons why it cannot do so. How law and society will grapple with unconscious cognitions is difficult to predict, even in the near future. Yet, a long range prediction is precisely what we attempt in our "future history."

[8] Part I introduces the science with a conceptual primer and summary of the scientific data. With this as background, Part II articulates our call for behavioral realism. Part III provides a stylized prediction of how we might reach a new scientific consensus, and how such a consensus might be received into the law. In our prognostications, we expect to see modes of argument and reaction that we label denial, minimization, politicization, resignation, and justification. Along the way, we revisit *Juan's Interview*, *Skip's Encounter*, and *Grace's Hire*, to see how the mind sciences simultaneously complicate and clarify those pictures.

² See Jerry Kang & Mahzarin Banaji, *Fair Measures: A Behavioral Realist Revision of Affirmative Action*, 94 CALIF. L. REV. 1063, 1064 (2006).

I. SCIENCE PRIMER

[9] Some “truths” resonate so strongly with personal experience that they could not possibly be wrong. Consider what parents simply *know* to be true: sugar causes hyperactivity in their children. After all, sugar is energy, and all parents have struggled with kids who won’t sleep after coming home stuffed with cake and soda. The problem is that multiple scientific studies have shown that this is just plain wrong.³ Why, then, does such a myth persist?

[10] We may be misled by a confound: the manic play at the birthday party—and not the sugar—may be the actual stimulant. Moreover, we suffer from confirmation biases. Once the “sugar causes hyperactivity” theory is in place, we pay attention and remember theory-confirming events but not theory-disconfirming ones.⁴ Beliefs about our own minds work similarly. The fact that thinking and feeling happen often without conscious awareness, control, or intention seems implausible. But like the conviction that sugar causes hyperactivity, what we simply *know* about ourselves may be wrong.

A. The Psychological Context

[11] Before focusing on implicit social cognition (ISC), we review the broader psychological context. First, it is well-settled that behavior is often caused by factors we are unaware of, and frequently can’t identify. There is substantial evidence, for example, of implicit memory, implicit perception (e.g., perception that happens without conscious awareness of the stimuli that nonetheless can influence behavior), and implicit learning (e.g., people can learn patterns even if they cannot report learning the sequence).⁵

[12] Sometimes, even exposure to trivial stimuli can alter behavior.⁶ For example, after a subtle reminder of the elderly, people remembered less⁷ and

³ See Mark L. Wolraich et al., *The Effect Of Sugar On Behavior Or Cognition In Children: A Meta-Analysis*, 27 JAMA 1617, 1621 (1995) (conducting a meta-analysis and finding that “sugar . . . does not affect the behavior or cognitive performance of children”).

⁴ See, e.g., Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon In Many Guises*, 2 REV. OF GENERAL PSYCHOLOGY 175 (1998) (literature review); Charles G. Lord et al., *Biased Assimilation And Attitude Polarization: The Effects Of Prior Theories On Subsequently Considered Evidence*, 37 J. PERSONALITY & SOCIAL PSYCHOLOGY 2098 (1979).

⁵ See, e.g., Daniel B. Willingham & Kelly Goedert-Eschmann, *The Relation Between Implicit And Explicit Learning: Evidence For Parallel Development*, 10 PSYCH. SCIENCE 531 (1999); Mary J. Nissen & Peter Bullemer, *Attentional Requirements Of Learning: Evidence From Performance Measures*, 19 COG. PSYCH. 1 (1987).

⁶ For a review, see S. Christian Wheeler & Richard E. Petty, *The Effects Of Stereotype Activation On Behavior: A Review Of Possible Mechanisms*, 127 PSYCH. BULL. 797 (2001).

⁷ See Ap Dijksterhuis et al., *On The Relation Between Associative Strength And Automatic Behavior*, 36 J. EXPERIMENTAL SOC. PSYCH. 531, 536-37 (2000) (finding that a subliminal prime caused decrements in memory performance among people with substantial contact with the elderly); Becca Levy, *Improving Memory in Old Age Through Implicit Self-Stereotyping*, 71 J. PERSONALITY & SOC. PSYCHOL. 1092, 1097-98 (1996) (finding that subliminal priming of negative stereotypes of aging worsened memory performance).

walked more slowly down a hallway.⁸ Exposure to a briefcase (versus backpack) led people to be more self-interested;⁹ the smell of Lysol led people to eat more neatly;¹⁰ holding a hot (versus cold) cup of coffee prompted warmer personality ratings of others.¹¹ Behavioral economists and decisionmaking psychologists have shown how behavior is not always a product of conscious and rational reflection, as demonstrated, for example, by the endowment effect¹² and the availability heuristic.¹³ These various findings share a common thread: navigation of the social world often happens automatically, and without conscious attention, reflection, or awareness.

[13] Second, even accurate identification of behavioral influences does not guarantee effective control of that behavior. One can experience this “lack of control” first-hand by taking the Stroop Test.¹⁴ In one portion of the task, words are written in the color they describe. The word “red” appears in red ink, and the word “blue” appears in blue ink and so on. Naming the color in which the words appear is trivial. However, if the words are changed so that the content of the word does not match the color of the ink (i.e., the word “red” is now printed in blue ink), naming the color becomes significantly harder. Comprehending the meaning of words is so automatic that the text prevents us from setting that information aside and doing what we consciously intend to do (that is, name the color).

[14] Third, introspection does not tell us much of our mind’s contents. Large bodies of cognitive and social psychology research make it untenable to assume that social behavior can be understood by attending only to what is available in conscious thought.¹⁵ As Richard Nisbett and Timothy Wilson reported back in

⁸ See, e.g., 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, 236-38 (1996) (demonstrating that activating the elderly stereotype altered the speed of walking)

⁹ See Aaron C. Kay et al., *Material Priming: The Influence Of Mundane Physical Objects On Situational Construal And Competitive Behavioral Choice*, 95 ORG. BEH. & HUM. DECISION PROCESSES 83, 89-91 (2004) (finding that participants retained more of the \$10 they were asked to split between themselves and a partner than they did when exposed to icons from the college world such as a backpack) (Studies 3 and 4, the “Ultimatum game”).

¹⁰ See Rob W. Holland et al., *Smells Like Clean Spirit: Nonconscious Effects Of Scent On Cognition And Behavior*, 16 PSYCHOLOGICAL SCIENCE 689, 691-92 (2005).

¹¹ See Lawrence E. Williams & John A. Bargh, *Experiencing Physical Warmth Promotes Interpersonal Warmth*, 322 SCIENCE 606 (2008) (Study 1).

¹² See generally Russell B. Korobkin, 97 NORTHWESTERN U. L. REV. 1227 (2003) (literature review).

¹³ See HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT 196-98 (THOMAS GILOVICH ET AL. EDS. 2002); RICHARD E. NISBETT & LEE ROSS, HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT (1980); Norbert L. Kerr et al., *Bias in judgment: Comparing individuals and groups*, 103 PSYCH. REV. 687 (1996); Norbert Schwarz et al., *Ease Of Retrieval As Information: Another Look At The Availability Heuristic*, 61 J. PERSONALITY & SOCIAL PSYCH. 195 (1991).

¹⁴ See J. Ridley Stroop, *Studies of interference in serial verbal reactions*, 18 J. EXPERIMENTAL PSYCHOLOGY 643 (1935).

¹⁵ See Richard E. Nisbett & Timothy DeCamp Wilson, *Telling More Than We Can Know: Verbal Reports On Mental Processes*, 84 PSYCH. BULL. 231, 243-44 (1977). An updated version of the argument

1977, “[t]he accuracy of subjective reports is so poor as to suggest that any introspective access that may exist is not sufficient to produce generally correct or reliable reports.”¹⁶ They acknowledged how disquieting this may be: “It is frightening to believe that one has no more certain knowledge of the working of one’s own mind than would an outsider with intimate knowledge of one’s history and of the stimuli present at the time the cognitive process occurred.”¹⁷ In the more than three decades since their landmark article, social and cognitive psychologists have repeatedly demonstrated this same simple fact: introspection as a way of knowing (and reporting on) one’s mind is severely limited.

B. The Challenge of Implicit Social Cognition

[15] With this context, we are now ready to define psychological concepts fundamental to implicit social cognition. First, psychologists distinguish “attitude” from “stereotype.” An *attitude* is “an association between a given object and a given evaluative category.”¹⁸ Evaluative categories are either positive or negative, and as such, attitudes reflect what we like and dislike, favor and disfavor, approach and avoid. For example, “I dislike Mexicans” is an example of an attitude, consciously reported. (A negative attitude is often called “prejudice.”) By contrast, a *stereotype* does not emphasize an evaluative association as much as a more specific attribute, such as “Mexicans are undocumented.” Of course, a particular stereotype may support an overall evaluative attitude. If one dislikes undocumented people and one associates Mexicans with undocumented status, then it is likely that this attribute will contribute toward a negative attitude toward Mexicans.

[16] Second, psychologists distinguish “implicit” from “explicit.” Similar to the usage in cognitive psychology, the term “implicit” emphasizes that we are unaware of our having a particular thought or feeling.¹⁹ Further, we might even reject that thought or feeling as inaccurate or inappropriate upon self-reflection. By contrast, “explicit” emphasizes awareness of having a thought or feeling; accordingly, we are able to self-report having such cognitions. Typically,

appears in Timothy D. Wilson & Elizabeth W. Dunn *Self-Knowledge: Its Limits, Value, And Potential For Improvement*, 55 ANNUAL REV. PSYCHOLOGY 493, 494 (2004) (discussing the “several reasons why people are not an open book to themselves”).

¹⁶ Nisbett & Wilson, *supra* note 15, at 233.

¹⁷ *Id.* at 257.

¹⁸ Russell H. Fazio et al., *Attitude accessibility, attitude-behavior consistency, and the strength of the object-evaluation association*, 18 J. EXPERIMENTAL SOCIAL PSYCHOLOGY 339, 341 (1982).

¹⁹ Anthony G. Greenwald & Mahzarin R. Banaji, *Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes*, 102 PSYCHOL. REV. 4, 8 (1995) (“Implicit attitudes are introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects.”). As regards to the “inaccurately identified” qualification, the authors explain that “a student may be aware of having been graded highly in a course, but not suspect that this experience influences responses to the course’s end-of-term course evaluation survey.” *Id.* at n.2

although not necessarily, we agree with and endorse our explicit thoughts and feelings.

[17] These basic concepts—attitudes and stereotypes, explicit and implicit—can be applied to objects, like bricks and broccoli. For example, we may associate bricks with houses even though most houses aren't made of bricks. In this way, we have a stereotype about houses. The existence of implicit cognitions about objects, such as houses, has been uncontroversial. But as applied to humans, these basic concepts have been sharply protested. To be sure, it is more unsettling to think that we may hold two separate (and often divergent) attitudes toward a social group, such as Mexicans, as compared to mere objects, such as broccoli. But our anxiety says nothing about whether implicit social cognitions in fact exist and influence our behavior. What legal and policy analysts must appreciate is that the findings in nonsocial and social domains rest on identical scientific foundations.

C. The Techniques

[18] Suppose we are eager to collect data about our implicit attitudes and stereotypes about social groups. How might we do so? Historically, psychologists, sociologists, and political scientists have measured mental content by asking people for self-reports. There are, however, two limitations to this methodology.²⁰ First, survey participants may engage in impression management, providing what they feel are politically correct answers.²¹ The second and deeper problem is that introspection is undependable.²²

[19] To overcome these difficulties, cognitive and social psychologists have developed instruments to measure implicit cognitions without asking for self-reports.²³ Some tests use linguistic cues;²⁴ others use physiological instru-

²⁰ See Jerry Kang, *Trojan Horses of Race*, 118 HARV. L. REV. 1491, 1506 (2005) (calling this the "opacity" problem).

²¹ Even without any conscious strategy to deceive, respondents might alter responses to align with perceived questioner expectations. Cf. Stacey Sinclair et al., *Social Tuning of Automatic Racial Attitudes: The Role of Affiliative Motivation*, 89 J. PERS. & SOC. PSYCH. 583, 590 (2005) (finding that "automatic prejudice shifted toward the ostensible attitudes of a social actor to the degree that individuals were motivated to get along with him or her").

²² See supra ¶ [14].

²³ See Russell H. Fazio & Michael A. Olson, *Implicit Measures In Social Cognition: Their Meaning And Use*, 54 ANNUAL REVIEW OF PSYCHOLOGY 297-327 (2003) (reviewing various implicit measures in general and semantic priming and the IAT in particular).

²⁴ See, e.g., Denise Sekaquaptewa et al., *Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination*, 39 J. EXPERIMENTAL SOCIAL PSYCHOLOGY 75 (2003); Lisa Sinclair & Ziva Kunda, *Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes*, 77 J. PERSONALITY & SOCIAL PSYCHOLOGY 885 (1999) (using word fragment completion method); William Von Hippel et al., *The Linguistic Intergroup Bias As an Implicit Indicator of Prejudice*, 33 J. EXPERIMENTAL SOC. PSYCHOL. 490, 507 (1997) (seeing whether mere description versus explanation is proffered to describe a behavior).

ments, for instance, by measuring cardiovascular responses,²⁵ micro-facial movements,²⁶ or neuron activity.²⁷ But the largest class of measures relies on reaction times when completing various tasks. The basic assumption in such tests is that mentally simple tasks take a (relatively) short time to complete, whereas mentally difficult tasks take a (relatively) long time to complete. For instance, any two concepts that are associated together in our minds will be easier to pair together in a timed task. After hearing the word “doctor,” the word “nurse” comes to mind much more quickly than the word “purse.” By measuring the speeds of activations, we can infer their strength of association. As simplistic as this approach sounds, reaction time instruments²⁸ have produced the most reliable and popular measures for implicit social cognitions.

[20] Among these instruments, the most widely used is the Implicit Association Test (IAT).²⁹ The IAT requires participants to classify rapidly individual stimuli

²⁵ See Jim Blascovich et al., *Perceiver threat in social interactions with stigmatized others*, 80 J. PERSONALITY & SOCIAL PSYCHOLOGY 253 (2001).

²⁶ See Eric J. Vanman et al., *Racial discrimination by low-prejudiced Whites facial movements as implicit measures of attitudes related to behavior*, 15 PSYCHOLOGICAL SCIENCE 711 (2004).

²⁷ See Jason P. Mitchell et al., *Thinking about Others: The Neural Substrates of Social Cognition*, in SOCIAL NEUROSCIENCE: PEOPLE THINKING ABOUT THINKING PEOPLE 63 (John T. Cacioppo et al., Eds. 2006); Elizabeth A. Phelps, et. al., *Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation*, 12 J. COG. NEUROSCIENCE 729, 729 (2000).

²⁸ There are many such instruments. One commonly-used reaction time measure is *semantic priming*, which was initially designed to measure memory association strengths. As explained, people respond more quickly to the word “nurse” after having recently been exposed to the word “doctor.” See David E. Meyer & Roger W. Schvaneveldt, *Facilitation recognizing pairs of words: Evidence of a dependence between retrieval operations*, 90 J. Experimental Psychology 227 (1971).

This design was later modified to study *evaluative priming*—the extent to which a concept is associated with the concepts “good” and “bad.” See Russell H. Fazio et al., *On the automatic activation of attitudes*, 50 J. PERSONALITY & SOCIAL PSYCHOLOGY 229 (1986). In this measure, pairs of items appear serially on a computer screen, and participants are asked to ignore the first item and respond to the second one. For example, in a task that assesses automatic attitudes, participants must press computer keys to categorize a word as either “good” or “bad.” See Russell H. Fazio et al., *Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline?*, 69 J. PERSONALITY & SOC. PSYCHOL. 1013, 1015-1016 (1995) (describing the first adaptation of the evaluative priming procedure to measure automatic racial attitudes). Each word is, however, preceded by a brief picture of, say, pizza or vomit. After seeing the vomit (compared to when they saw the pizza), people were slower to respond to good words such as “happy” or “sunshine” and faster to respond to bad words such as “awful” or “terrible.” The inference is that, on average, the vomit activated negativity more than did the pizza. Interestingly, we can switch from pizza and vomit to pictures of Whites and African Americans, and see similar results. See *id.*

²⁹ Seminal work in this domain was performed by Fazio who explored how “priming” could help measure automatic attitudes and beliefs. See B. Wittenbrink, *Measuring attitudes through priming*, in IMPLICIT MEASURES OF ATTITUDES 17 (B. Wittenbrink & N. Schwarz, Eds. 2007). Newer tasks, such as the Go/ No-Go Association Task, see Brian A. Nosek & Mahzarin A. Banaji, *The Go/No-go Association Task*, 19 SOCIAL COGNITION 625 (2001), the Evaluative Movement Assessment, see C.M. Brendl et al., *Indirectly measuring evaluations of several attitude objects in relation to a neutral reference point*, 41 J. EXPERIMENTAL SOCIAL PSYCHOLOGY 346 (2005), the extrinsic affective Simon task, see J. De Houwer, *The extrinsic affective Simon task*, 50 Experimental Psychology 77 (2003), and the affect misattribution procedure, see B.K. Payne, *An inkblot for attitudes: Affect misattribution as implicit measurement*, 89 J. PERSONALITY & SOC. PSYCHOL. 277 (2005), have been used, and likely more will be developed.

into one of four distinct categories using only two responses (for example, in the computerized IAT, participants might respond using only the “E” key on the left side of the keyboard, or “I” on the right side).³⁰ For instance, in a race attitude IAT, there are two social categories, WHITE and ASIAN, and two attitudinal categories, GOOD and BAD. WHITE and ASIAN might be represented by black-and-white photographs of the faces of White and Asian people. GOOD and BAD could be represented by words that are easily identified as being linked to positive or negative affect, such as “joy” or “agony”.³¹ A person with a negative implicit attitude toward ASIAN would be expected to go more quickly when ASIAN and BAD share one key, and WHITE and GOOD the other, than when the pairings of good and bad are switched.³² The time difference, which is recalibrated into what is known as the IAT D score, is interpreted as reflecting an implicit attitude.³³

D. The Findings

[21] For brevity’s sake, in this section, we focus on the Implicit Association Test (IAT) although other instruments have produced comparable findings. Since 1998 when the IAT was officially introduced, hundreds³⁴ of peer-reviewed scientific publications have produced largely consistent results. First, implicit biases are both pervasive (large numbers of individuals show evidence of them), and large, statistically speaking. Second, implicit biases show dissociation from explicit biases, which are the consciously reported preferences (attitudes) and beliefs (stereotypes) of the same individual. Third, implicit biases predict behavior. Fourth, despite their seemingly uncontrollable nature, implicit biases are malleable and highly responsive to environmental cues.

³⁰ See Anthony G. Greenwald et al., *Measuring individual differences in implicit cognition: The Implicit Association Test*, 74 J. PERSONALITY & SOC. PSYCHOL. 1464 (1998) (introducing the IAT).

³¹ In the standard attitude IAT, Good is represented by “Joy, Love, Peace, Wonderful, Pleasure, Friend, Laughter, Happy.” Bad is represented by “Agony, Terrible, Horrible, Nasty, Evil, War, Awful, Failure.” Before 1999, Project Implicit used the term “Death” instead of “Horrible.” See Brian A. Nosek, Mahzarin R. Banaji & Anthony G. Greenwald, *Harvesting Implicit Group Attitudes and Beliefs from a Demonstration Web Site*, 6 GROUP DYNAMICS 101, 114 (2002).

³² For further information on the IAT, see Kristin A. Lane et al., *Understanding and using the Implicit Association Test: IV: What we know (so far) about the method*, in IMPLICIT MEASURES OF ATTITUDES 17 (B. Wittenbrink & N. Schwarz, Eds. 2007); Brian Nosek, Anthony Greenwald, Mahzarin R. Banaji, *The Implicit Association test at Age 7: A methodological and Conceptual Review*, in AUTOMATIC PROCESSES IN SOCIAL THINKING AND BEHAVIOR 265 (JOHN A. BARGH, ED. 2007).

³³ See Jerry Kang et al., *Are Ideal Litigators White? Measuring the Myth of Colorblindness*, ¶ 57 (July 31, 2009) (<http://ssrn.com/abstract=1442119>) (finding negative implicit attitude against Asian Americans) (AT D = 0.62), $t(67) = 13.31$, $p < .001$

³⁴ A PsycINFO database search for “implicit association test” anywhere in a record produced 1,725 separate publications, in 1353 peer-reviewed journals. If the field was restricted just to keywords, the search produced 449 entries, with 380 in peer-reviewed journals. See screen capture, 2/16/2009 12:31 PM (on file with author).

1. Pervasiveness

[22] Psychologists have long documented *ingroup favoritism*³⁵—our tendency to favor the groups we belong to. Implicit measures also register such preferences. For example, on IATs, Japanese-Americans and Korean-Americans preferred their own ethnic group relative to the other,³⁶ as did East and West Germans.³⁷ Ingroup bias is so strong that people explicitly report liking “ingroups” even when they are randomly assigned to them,³⁸ and even when the groups are made up. For example, being arbitrarily placed in the category ‘Quan’ or ‘Xanthie’—groups that do not exist, and thus should be considered equivalent—generated implicit biases within participants in favor of their assigned group.³⁹

[23] However, in the real world, social groups are rarely equivalent; instead, they exist in hierarchies. What impact might these social hierarchies have on the otherwise unremarkable tendency to prefer one’s ingroup? Clear evidence comes from the data collected through Project Implicit,⁴⁰ a research website operated by Harvard University, Washington University, and the University of Virginia. At Project Implicit, visitors can try IATs that examine both implicit attitudes and stereotypes on topics ranging across race, gender, age, politics, region, religion, even consumer brands. With over 7 million tests completed, this is the largest repository of implicit social cognition data.⁴¹

[24] Table 1 provides results from the seventeen IATs available at Project Implicit.⁴² Most participants demonstrated implicit attitudes in favor of one social group over another, away from the neutral position of no bias. Put

³⁵ See Henri Tajfel & John C. Turner, *The Social Identity Theory of Intergroup Behavior*, in *POLITICAL PSYCHOLOGY: KEY READINGS IN SOCIAL PSYCHOLOGY* 276 (JOHN T. JOST & JIM SIDANIUS, EDS. 2004).

³⁶ See Greenwald, *supra* note 30, at 1471-1472.

³⁷ See Ulrich Kuhnen et al., *How robust is the IAT? Measuring and manipulating implicit attitudes of East- and West-Germans*, 48 *ZEITSCHRIFT FÜR EXPERIMENTELLE PSYCHOLOGIE* 135 (2001).

³⁸ See Maria Rosaria. Cadinu & Myron. Rothbart M, Self-anchoring and differentiation processes in the minimal group setting, 70 *J. PERSONALITY & SOC. PSYCHOL.* 661 (1996); Samuel L. Gaertner et al., Reducing intergroup bias: The benefits of recategorization, 57 *J. PERSONALITY & SOC. PSYCHOL.* 239 (1989); Henri Tajfel et al., *Social categorization and intergroup behavior*, 1 *EUROPEAN J. SOCIAL PSYCHOLOGY* 149 (1971).

³⁹ See L. Ashburn-Nardo et al., *Implicit associations as the seeds of intergroup bias: How easily do they take root?*, 81 *J. PERSONALITY & SOC. PSYCHOL.* 789 (2001).

⁴⁰ <<http://projectimplicit.org>>

⁴¹ Because data are gathered from volunteers, they do not reflect a random sample of the population. However, this sample is more demographically representative than the narrow pool typically used in economics and psychology experiments (often college student volunteers). In any event, data from web-based sample (from volunteers across the globe) converge with laboratory data. Additionally, the enormous size of the data repository from the Internet harvest allow questions to be answered with confidence that is not otherwise possible. For general discussion about the benefits and costs of using Internet data, see Nosek et al., *Harvesting*, *supra* note 31, at 102-04.

⁴² See Brian A. Nosek et al., *Pervasiveness and Correlates of Implicit Attitudes and Stereotypes*, 18 *EURO. REV. SOCIAL PSYCH.* 1 (2007).

another way, notwithstanding protestations to the contrary, people are generally not blind to race, gender, religion, social class, or other demographic characteristics. More important, participants systematically preferred socially-privileged groups: Young over Old, White over Black, Light Skinned over Dark Skinned, Other Peoples over Arab/Muslim, Abled over Disabled, Thin over Obese, and Straight over Gay.⁴³

[25] [insert Table 1 about here]

[26] These hierarchy-driven biases interact with ingroup biases. Accordingly, those who belong to social groups deemed to be “good” (e.g., the young, European Americans, straight) show strong preference for their own group. On the other hand, those who come from groups the culture assigns as “bad” (e.g., elderly, African Americans, gay) do not show strong ingroup preference. Often their data show no ingroup preference at all and sometimes even tilt in the opposite direction.⁴⁴ In laboratory studies, we see similar findings for participants who belong to lower-status social groups marked by weight, age, sexual orientation, socioeconomic class,⁴⁵ university status (i.e. students at lower-ranked universities),⁴⁶ and college dorm status (i.e. students at less popular dorms).⁴⁷ In other words, the overweight, poor, and those associated with less-valued institutions all showed weaker preference for their own group on implicit measures.

[27] Let us now shift focus away from implicit attitudes to implicit *stereotypes*. Recall that stereotypes address the likelihood that a social category features some attribute. So, for instance, the question isn’t whether one likes or dislikes men versus women (an attitude);⁴⁸ instead, the question becomes whether men are more associated with attributes such as tall, career, or mathematics whereas women are more associated with short, family, or arts. Again, the data show that implicit stereotypes are pervasive and robust. For instance, most partici-

⁴³ See *id.* at 11.

⁴⁴ See Leslie Ashburn-Nardo et al., *Black Americans' implicit racial associations and their implications for intergroup judgment*, 21 *SOCIAL COGNITION* 61, 71 (2003); Robert W. Livingston, *The role of perceived negativity in the moderation of African Americans' implicit and explicit racial attitudes*, 38 *J. EXPERIMENTAL SOCIAL PSYCHOLOGY* 405 (2002). Africans in the U.S. and in South Africa show substantially weaker ingroup preference. See Kristin R. Olson et al., *Implicit intergroup attitudes in South Africa*. Poster presented at the 9th Annual Meeting of the Society for Personality and Social Psychology, Albuquerque, NM (2008).

⁴⁵ See Laurie A. Rudman et al., *Minority members' implicit attitudes: Automatic ingroup bias as a function of group status*, 20 *SOCIAL COGNITION* 294 (2002).

⁴⁶ See John T. Jost et al., *Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups*, 38 *J. EXPERIMENTAL SOCIAL PSYCH.* 586 (2002).

⁴⁷ See Kristin A. Lane et al., *Me and my group: Cultural status can disrupt cognitive consistency*, 23 *SOCIAL COGNITION* 353 (2005).

⁴⁸ In an important counter-example of ingroup favoritism, “men are less likely than women to show automatic ingroup bias (i.e., own gender preference).” See Laurie A. Rudman & Stephanie A. Goodwin, *Gender Differences in Automatic In-Group Bias: Why Do Women Like Women More Than Men Like Men?*, 87 *J. PERSONALITY & SOC. PSYCHOL.* 494, 494 (2004). A positive attitude toward women doesn’t say anything about whether we stereotype women.

pants (72%) associated the concepts MALE with SCIENCE and FEMALE with HUMANITIES.⁴⁹ Similarly, participants found it easier to categorize WHITE, rather than ASIAN AMERICAN or NATIVE AMERICAN, faces with AMERICAN, reflecting an implicit stereotype that “American = White.”⁵⁰

2. Dissociation

[28] Scores on implicit measures do not map neatly to explicit self-reports provided on surveys. This gap is called “dissociation.” As Table 1 shows, explicit biases are typically far smaller than their implicit counterparts, as measured in standard units.⁵¹

[29] It would be odd to conclude that because implicit bias measures differ from explicit bias measures that the former are spurious. The entire point was to unearth attitudes and stereotypes that cannot be explicitly recalled or accurately expressed. At the same time, it would be presumptuous to conclude that all explicit bias measures are worthless because implicit biases are the only “true” measures.⁵² Most scientists in the field reject such privileging, as do we.⁵³ The best understanding is that implicit and explicit measures tap separate but related sets of cognitions.⁵⁴

⁴⁹ Nosek et al., *supra* note 42, at 11 (table 2) ($M=.37$; $d=.93$).

⁵⁰ See Thierry Devos & Mahzarin R. Banaji, *American = White?*, 88 J. PERSONALITY & SOC. PSYCHOL. 447, 453, 457 (2005).

⁵¹ A standard unit measures magnitude in terms of its number of standard deviations. These data come from Nosek et al., *supra* note 42.

⁵² See Fazio & Olson, *supra* note 23; Lane et al., *supra* note 31.

⁵³ Further research on dissociation reveals that the correlations between implicit and explicit measures vary across study, target group, and participant characteristics. One way to analyze such varied results is to conduct a meta-analysis, which quantitatively synthesizes all the studies on a particular topic. Hofmann and colleagues’ meta-analysis of 126 correlations between implicit (assessed with the IAT) and explicit attitudes revealed a mean population correlation $r=0.24$, which represents a moderate strength relationship between the two variables. See Wilhelm Hofmann, *A Meta-Analysis on the Correlation Between the Implicit Association Test and Explicit Self-Report Measures*, 31 PERSONALITY & SOC. PSYCH. BULL. 1369 (2005). That is, explicit and implicit biases are somewhat related, but not highly so.

⁵⁴ See Greenwald & Banaji, *supra* note 19; Timothy D. Wilson, *A model of dual attitudes*, 107 PSYCHOLOGICAL REVIEW 101 (2000). A statistical procedure known as confirmatory factor analysis (CFA) checks whether the explicit surveys and the IAT scores tap the same underlying mental construct, in which case they would, in CFA’s language, load onto a single factor (indicating that they are two different ways of measuring the same psychological trait). Across numerous possible attitude objects (race, age, gender, movie stars, favorite foods), implicit and explicit measures appear to be separate but related mental constructs. This pattern held true for 56 out of 57 different attitude objects. See Brian A. Nosek, *Moderators of the relationship between implicit and explicit evaluation*, 134 J. EXPERIMENTAL PSYCHOLOGY: GENERAL 565 (2005).

3. Real-world Behaviors

[30] All that we have shown so far are pervasive differences in some computerized test. But do these differences, measured in milliseconds, matter in the real world?

[31] *Juan's Interview*. Consider, for example, why Juan's interview might have gone badly. Of course, it's impossible to know for sure in any particular case. Still, we have learned that implicit attitudes correlate with facial expressions, eye contact, and body posture.⁵⁵ In one of the earliest demonstrations of implicit racial attitudes, people with more negative implicit attitudes toward Blacks (compared with Whites) on an evaluative priming measure were rated less friendly by Black evaluators.⁵⁶ This unfriendliness seeps out through nonverbal cues: the more negative the implicit attitude, the more awkward the body language.

[32] In fact, even the same facial expression can be interpreted differently as a function of implicit bias. In one study, participants watched a video of computer-generated faces that morphed slowly from a frown to a smile and were instructed to hit a key when they thought the expression changed. In general, people saw hostility "linger" on the Black face for a longer period of time. Moreover, the extent that hostility was perceived as lingering was predicted by implicit bias (as measured by the IAT) against Blacks.⁵⁷

[33] Body language matters because social psychologists have long confirmed that if one person acts awkwardly to another, the other person will reciprocate thus generating a vicious circle.⁵⁸ That interaction will then sour the interview without either party recognizing the implicit causal forces. *If Juan's interviewer*

⁵⁵ See Kristi M. Lemm, *Positive associations among interpersonal contact, motivation, and implicit and explicit attitudes toward gay men*, 51 J. HOMOSEXUALITY 79 (2006); Allen R. McConnell & Jill M. Leibold, *Relations Among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes*, 37 J. EXPERIMENTAL SOC. PSYCHOL. 435, 435-36 (2001).

⁵⁶ See Fazio et al., *supra* note 28.

⁵⁷ See Kurt Hugenberg & Galen V. Bodenhausen, *Facing prejudice: Implicit prejudice and the perception of facial threat*, 14 PSYCH. SCI. 640, 641 (2003) (showing that when the video went from hostile to friendly, implicit anti-Black bias predicted the extent to which hostility would "linger"; $p = .04$). Implicit bias scores predicted responses to Black $\beta = .46$, $p = .02$, but not White, $\beta = .09$, *ns*, faces. See *id.* at 642. In a second study, the video showed faces going from friendly to hostile. In this situation, people higher in implicit bias were quicker to report detecting hostility in Black rather than White faces, $p = .02$. See *id.* at 642.

⁵⁸ See Carl O. Word et al., *The Nonverbal Mediation of Self-Fulfilling Prophecies in Interracial Interaction*, 10 J. EXPERIMENTAL SOC. PSYCHOL. 109 (1974) (finding that when White interviewers treated White interviewees with unfriendly nonverbal behavior, the White interviewees gave worse interviews as measured by third party evaluators blind to the purpose of the experiment); Mark Chen & John A. Bargh, *Nonconscious Behavioral Confirmation Processes: The Self-Fulfilling Consequences of Automatic Stereotype Activation*, 33 J. EXPERIMENTAL SOC. PSYCHOL. 541, 554-55 (1997) (showing that hostility in one game partner, induced by racial priming, induced hostility in the other game partner in a "password" game).

had a negative implicit attitude toward Mexicans, that attitude may have leaked through his body language, which could have triggered a downward spiral.⁵⁹

[34] *Skip's Encounter*. If “doctor” brings “nurse” to mind, could the sight of a Black man brings “criminal” to mind? If a neutral facial expression is seen as more angry on a Black face, could a wallet be mistaken for a weapon more often when held by a Black man? To both questions, the data suggest “yes.”

[35] A “shooter bias” paradigm has explored whether people are more prone to accidentally shoot Black than White suspects. In this task, individuals (targets) holding an object appear in front of ordinary background scenes such as bus stations and parks. Participants view these scenes on a computer and have a simple task: make one response if the person holds a weapon, and another response if the person holds a harmless object such as a cell phone or wallet.⁶⁰

[36] Responses differed as a function of the target’s race: Participants were quicker to “shoot” an armed Black target than an armed White target, but slower to “not shoot” an unarmed Black target than an unarmed White target.⁶¹ When time-pressured (to better mimic life-or-death decisions), unarmed Black targets were mistakenly shot more often than unarmed White targets, and armed White targets were mistakenly not shot more often than unarmed Black targets.⁶² Interestingly, these stereotype-consistent behaviors emerged among both Black and White participants.⁶³ On the margins, then, *Skip's Encounter* was more dangerous because of his race, regardless of the cop’s race.

[37] *Grace's Hire*. Faculty supporters claimed that Grace is a “role model” who is counter to the stereotypes associated with Asian American women. They claim that adding her to the faculty would do “good” for faculty as well as the student body. Plausible and familiar claims, but what are the data?

[38] Well, it turns out that there is some support for the idea. First, imagining a counter-typical female leader (as opposed to a Caribbean vacation) for a few

⁵⁹ Other research, however, suggests that, under certain circumstances, these attitudes may leak out in the opposite way. During a brief race-related conversation, Black partners actually liked Whites with *higher* levels of implicit racial bias (as measured by the IAT) more than those with lower levels of bias. This disparity was accounted for by differences in perceived levels of engagement – Black partners found Whites with higher levels of implicit bias to be more engaged in the interaction. The authors speculated that during a race-salient conversation, Whites with higher bias may have been more active in trying to mask bias. J. Nicole Shelton et al., *Ironic effects of racial bias*, 16 PSYCHOLOGICAL SCIENCE 397 (2005).

⁶⁰ Joshua Correll et al., *The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals*, 83 J. PERSON. & SOC. PSYCH. 1314 (2002). 1315-1317 describing procedure. See Anthony G. Greenwald et al. *Targets of discrimination: Effects of race on responses to weapons holders*, 39 J. OF EXPERIMENTAL SOC. PSYCH. 399 (similar results).

⁶¹ See Correll et al., *supra* note 60, at 1317

⁶² See *id.* at 1319

⁶³ See *id.* at 1324-1325

minutes in a mental imagery exercise reduced implicit gender stereotyping.⁶⁴ Second, exposure to positive exemplars of the disfavored social category (pictures of Tiger Woods and Martin Luther King) decreased implicit bias against that category significantly in magnitude and over a 24-hour period.⁶⁵ Finally, one longitudinal study found that women who attended a single-sex university for one year had their average group implicit stereotypes against women decrease to zero. By contrast, a control group of women who attended a co-ed university for one year had its average group implicit bias increase. What explained these movements? After examining various university environmental variables, such as course work and extracurriculars, Nilanjana Dasgupta and Shaki Asgari determined that the mechanism was exposure to female professors and administrators.⁶⁶ Collectively, these studies suggest not so much that Grace would be a “role model” but that she could function as a “debiasing agent,”⁶⁷ whose countertypical performance decreases implicit bias in the student body.⁶⁸

II. BEHAVIORAL REALISM

[39] The studies we have reviewed make a *prima facie* case for the existence of implicit social cognitions and their influence of real-world behavior. At the least, the science enriches our understanding of the stories of Juan, Skip, and Grace. Suppose that the supporting evidence accumulates even further into a widespread scientific consensus? What impact should this consensus have on the law and legal institutions?

[40] In approaching this question, we seek “behavioral realism” in the law. Behavioral realism involves a three step process:

First, identify advances in the mind and behavioral sciences that provide a more accurate model of human cognition and behavior.

Second, compare that new model with the latent theories of human behavior and decision-making embedded within the law. These latent

⁶⁴ Irene V. Blair, Jennifer E. Ma, & Alison P. Lenton, *Imagining Stereotypes Away: The Moderation of Implicit Stereotypes Through Mental Imagery*, 81 J. PERSON. & SOC. PSYCH. 828, 828 (2001).

⁶⁵ Nilanjana Dasgupta & Anthony G. Greenwald, *On the Malleability of Automatic Attitudes: Combating Automatic Prejudice with Images of Admired and the Disliked Individuals*, 81 J. PERSON. & SOC. PSYCH. 800, 802 (2001); Mitchell et al., *infra* note 80.

⁶⁶ See Nilanjana Dasgupta & Shaki Asgari, *Seeing is Believing: Exposure to Counterstereotypic Women Leaders and its Effect on the Malleability of Automatic Gender Stereotyping*, 40 J. EXPERIM. SOC. PSYCH. 642, 647 (2004).

⁶⁷ See Christine Jolls & Cass R. Sunstein, *The Law of Implicit Bias*, 94 CALIF. L. REV. 969 (2006) (discussing debiasing strategies).

⁶⁸ As compared to the “role model” justification, which courts have rejected, this debiasing agent justification may better survive legal challenge. See Kang & Banaji, *supra* note 2, at 1110-11. Under an Equal Protection analysis, for example, preventing racial discrimination is clearly a “compelling interest,” and the debiasing strategy may be sufficiently “narrowly tailored.” Crediting a candidate’s potential to be a “debiasing agent” in the overall assessment of her merit raises numerous difficult empirical, philosophical, ethical, and legal questions—all of which warrant cautious examination.

theories typically reflect “common sense” based on naïve psychological theories.

Third, when the new model and the latent theories are discrepant, ask lawmakers and legal institutions to account for this disparity. An accounting requires either altering the law to comport with more accurate models of thinking and behavior or providing a transparent explanation of “the prudential, economic, political, or religious reasons for retaining a less accurate and outdated view.”⁶⁹

[41] Although simple-sounding, this algorithm involves some complexity. For example, the *first* step raises the hard question of what counts as a “more accurate model” of human action. Whenever we have rapid advancements in instrumentation, measurement, and experimentation, we will have fractures within scientific understandings. Some scientists will cling to the traditional view; some will prefer the new. All sciences have periods of vigorous debate.⁷⁰ How certain, then, must a behavioral realist be before calling the new view more “accurate”? Should we adopt, for instance, a “precautionary principle”?⁷¹

[42] The *second* step requires identifying what folk-psychology is embedded within the status quo doctrine. But it is difficult to excavate the assumptions of human decisionmaking that have become “common sense.” Unmasking is challenging.

[43] The *third* step, which demands an “accounting,” is arguably the most complex. The strength of that demand depends “on numerous factors, such as the strength of the scientific consensus regarding the emergent model, the size of the gap between the new model and old assumptions, and the consequences of both action and omission.”⁷² In thinking about these consequences, one’s values inevitably come into play, and the behavioral realism approach itself says little about what those specific values should be.

[44] In this sense, behavioral realism reflects only weakly normative commitments. It reflects, first, a commitment to *reality over fantasy*. For instance, if the law assumed that the earth was flat and punished a teacher for conveying falsehoods, the fact that the earth is discovered *not* to be flat should prompt reformist pressure on the law. Fantasies about flat worlds—to the extent that they are demonstrated as fantasies—should bow to reality.

⁶⁹ Kristin Lane, Jerry Kang, & Mahzarin Banaji, *Implicit Social Cognition and the Law*, 3 ANNU. REV. LAW SOC. SCI. 19.1-19.25 (2007) [hereinafter “Annual Review”].

⁷⁰ For example, cancer researchers and advocates debated for decades about the effectiveness and utility of mammogram screenings for breast cancer, with the policy recommendations shifting to reflect the state of the science. See Donald A. Berry et al., *Effect of screening and adjuvant therapy on mortality from breast cancer*, 353 NEW ENGLAND J. MEDICINE 1784, 1792; Gina Kolata, *Mammograms validated as key in cancer fight*, NEW YORK TIMES, October 27, 2005.

⁷¹ See, e.g., David A. Dana, *A Behavioral Economic Defense of the Precautionary Principle*, 97 NW. U. L. REV. 1315, 1315 (2003) (discussing popularity of principle in environmental law and policy contexts).

⁷² See Lane, Kang, & Banaji, *Annual Review*, *supra* note 69, at 19.14-15.

[45] A second normative commitment is “*against hypocrisy and self-deception*.”⁷³ The law implicitly adopts some folk-psychology model of human behavior and decisionmaking in order to apportion responsibility and incentivize behaviors. But “garbage in” (i.e., incorrect models of the mind) will produce “garbage out” (unfair and inefficient rules and policies). New and better inputs should therefore produce new and better outputs. This is what the Supreme Court formally suggested in footnote 11 of its *Brown v. Board of Education* opinion, when it flagged “modern authority” as undermining the foundations of *Plessy*’s legacy of “separate but equal.”⁷⁴ That modern authority was a string cite to the psychological discoveries on the harm to Black children caused by segregation.⁷⁵ Notwithstanding compelling new evidence, the status quo must sometimes prevail. But the countervailing reasons should be spelled out and publicly defended; otherwise, we risk hypocrisy or self-deception in our law.

III. THE FUTURE HISTORY

A. Denial

[46] What drives any behavioral realist analysis is the novel, emerging scientific understanding. But maybe that understanding is just a horrible technical mistake, caused by instrumentation or mathematical error? This mode of response we label *denial*.

1. Scientific Denial

[47] Good scientists instinctively respond to any new measurement device with skepticism, and reaction time measures such as the IAT have been no exception.⁷⁶ In particular, scientists have carefully examined both its reliability and validity. By *reliability*, scientists mean that the instrument generates sufficiently reproducible measures over time. For example, a bathroom scale that gave radically different numbers each time you stepped on might be insufficiently reliable for someone trying to lose 5 pounds. By *validity*, scientists mean various things, among them statistical conclusion validity (did you run the statistics correctly?), internal validity (for any causal claims, are you sure there aren’t confounds), and construct validity (are you sure you’re actually measuring what you think you’re measuring?).

⁷³ Kang & Banaji, *Fair Measures*, *supra* note 2, at 1065.

⁷⁴ “Whatever may have been the extent of psychological knowledge at the time of *Plessy v. Ferguson*, this finding is amply supported by *modern authority*.” *Brown v. Bd. of Educ.*, 347 U.S. 483, 494 (1954) (emphasis added). We recognize that we are reading the Court at face value. That is why we say “formally suggested.”

⁷⁵ See *id.* n.11 (citing, e.g., the “doll studies” of K. B. Clark).

⁷⁶ To keep our discussion manageable, we again focus mostly on a single instrument, the IAT, although similar challenges and responses can be made of other measurement devices, ranging from the neurophysiologic to other reaction time measures.

[48] After a decade of research, we believe that the IAT has demonstrated enough reliability and validity such that total “denial” is implausible. As for reliability, across 20 studies, the average (and median) correlation between a person’s IAT score at two different times was 0.50,⁷⁷ which is respectable in terms of psychometric measures.⁷⁸ Because of the moderate reliability, nearly all scientists have discouraged using the IAT in high-stakes individual selection contexts, such as judicial nominations. But this level of reliability is not troubling for the IAT’s efficacy as a research tool because we can aggregate across people, to discern general patterns between mental constructs and behavior. One final counter-intuitive point: lower reliability levels (by adding noise) result in *underestimation* of causal relationships; so to the extent that we err, we do so by understating the impact of implicit bias.

[49] As for validity, questions about statistical conclusion validity have been addressed.⁷⁹ So have potential confounds that would undermine internal validity, sounding in terms of stimuli selection,⁸⁰ familiarity of the stimuli,⁸¹ the

⁷⁷ See Kristin A. Lane et al., *Understanding and using the Implicit Association Test: IV: What we know (so far) about the method*, in IMPLICIT MEASURES OF ATTITUDES 17, 70 (B. Wittenbrink & N. Schwarz, Eds. 2007) (aggregating test-retest reliabilities across 20 such administrations).

⁷⁸ This means that IAT scores at two time points share approximately 25% of their variance (.50-squared). A person’s score fluctuates around some mean.

⁷⁹ Statistical conclusion validity asks whether the numerical scores generated by instruments, such as the IAT, and any correlations have been analyzed correctly. For example, Blanton and Jaccard have questioned the meaningfulness of the IAT’s scale and challenged whether the IAT is a “ratio” scale; that is, whether a zero on the IAT reflects absence of bias. See H. Blanton & J. Jaccard, *Arbitrary metrics in psychology*, 61 AMERICAN PSYCHOLOGIST 27 (2006). This issue is hardly specific to the IAT. Unlike money in our pockets that dwindles down to nothing, it is extraordinarily difficult to determine when there is a total lack of a psychological trait such as “self-esteem,” “happiness,” or “racial preference.” For a response, see Anthony G. Greenwald, *Consequential Validity of the Implicit Association Test: Comment on Blanton and Jaccard*, 61 AMERICAN PSYCHOLOGIST 56 (2006) (noting that theories that predict multiplicative balanced relationships among cognitions (e.g., If I am an American and I am good, then Americans must be good) have been borne out by data and showing that the zero point on a Bush-Kerry IAT corresponded to the zero-point on the explicit measure, suggesting that both measures acted as ratio scales).

⁸⁰ See, e.g., Jason P. Mitchell et al., *Contextual variations in implicit evaluation.*, 132 J. EXPERIMENTAL PSYCHOLOGY: GENERAL 455 (2002) (finding that when items changed the categories’ construal, implicit biases were diminished). However, analyses of large web-based data sets suggest that when items do not change the construal of the category (i.e., the use of Kobe Bryant, Barry Bonds and Ronde Barber to represent the category “Black” could create the category “Athlete”) differences among them do not have much influence on IAT scores. Brian A. Nosek et al., *Understanding and using the Implicit and Association Test II: Method variables and construct validity* 31 PERSONALITY AND SOCIAL PSYCHOLOGY BULLETIN 166 (2005).

⁸¹ See Nilanjana Dasgupta et al., *The first ontological challenge to the IAT: Attitude or mere familiarity?*, 14 PSYCHOLOGICAL INQUIRY 238 (2003). Dasgupta et al. argued that three lines of evidence have “decisively laid the familiarity explanation to rest” at the *micro* level (the extent to which some of the items in the categories are more or less known to participants). First, controlling for familiarity did not influence results. Second, implicit preference for Whites emerged even when the task used pictures of people unknown to participants. Finally, preference for Whites over Blacks emerges even when the stimuli are carefully matched on frequency according to census data. See also SA Ottaway et al., *Implicit attitudes and racism: Effects of word familiarity and frequency on the implicit association test*, 19

order in which the combined tasks are completed,⁸² previous experience with the IAT,⁸³ inter-trial interval duration,⁸⁴ assignment of categories to right or left key,⁸⁵ right or left-handedness of participant,⁸⁶ cognitive fluency of participant,⁸⁷ and fakeability.⁸⁸ Finally, construct validity⁸⁹ challenges suggesting that

SOCIAL COGNITION 97 (2001); Nilanjana Dasgupta et al., *Automatic preference for White Americans: Eliminating the familiarity explanation*, 36 J. EXPERIMENTAL SOCIAL PSYCHOLOGY 316 (2000).

At the macro level, it seems unlikely that familiarity can account for the entirety of IAT effects because preferences emerge for arbitrary groups with nonsense names (such as “Xanthie”) to which participants are introduced and assigned to in the laboratory. See Leslie Ashburn-Nardo et al. *Implicit associations as the seeds of intergroup bias: How easily do they take root?*, 85 J. PERSONALITY & SOCIAL PSYCH 789.

⁸² See Greenwald et al., *supra* note 30. Brian A. Nosek et al., *Understanding and using the Implicit Association Test: II. Method variables and construct validity*, 31 PERSONALITY AND SOCIAL PSYCHOLOGY BULLETIN 166 (2005) (finding a small effect of order, which can be minimized by increasing the number of practice trials).

⁸³ See Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: 1. An Improved Scoring Algorithm*, 85 J. PERSONALITY & SOC. PSYCHOL. 197 (2003) (finding that prior experience slightly reduced the magnitude of IAT effects in subsequent trials).

⁸⁴ See Greenwald et al., *supra* note 30 (finding essentially no effect on the magnitude size of the IAT depending on the time interval, ranging from 100 ms to 700 ms, between successive trials of the IAT).

⁸⁵ See *id.* (finding no effects). The IAT counterbalances the side to which, for instance, pleasant and unpleasant categories are assigned.

⁸⁶ See Greenwald & Nosek, *supra* note 81 (finding no effects on self-reported handedness in large web-based data set).

⁸⁷ See Sam G. McFarland, & Zachary Crouch. *A cognitive skill confound on the Implicit Association Test*, 20 Social Cognition 483 (2002). Jan Mierke, & Karl C. Klauer. *Method-specific variance in the Implicit Association Test*, 85 J. Personality and Social Psychology 1180 (2003). Using improved scoring techniques reduces the unwanted influence of participants’ cognitive fluency. Huajian Cai et al. *The Implicit Association Test’s D measure can minimize a cognitive skill confound: Comment on McFarland and Crouch* (2002), 22 Social Cognition 673 (2004).

⁸⁸ See, e.g., Klaus Fiedler & Matthias Bluemke *Faking the IAT: Aided and unaided response control on the Implicit Association Tests*, 27 BASIC AND APPLIED SOCIAL PSYCHOLOGY 307 (2005); Melanie C. Steffens, *Is the Implicit Association Test immune to faking?* 51 EXPERIMENTAL PSYCHOLOGY 165 (2004); Do-Yeong Kim, *Voluntary controllability of the Implicit Association Test (IAT)*, 66 SOCIAL PSYCHOLOGY QUARTERLY 83 (2003).

These studies vary in their conclusions about the extent to which participants are able to control (or fake) their scores on the IAT. The best strategy to fake one’s IAT scores is to slow down intentionally on the compatible block (e.g., the *flower + good* pairing in an IAT measuring flower-insect attitudes). Although few subjects derive this strategy on their own, some certainly may. We note, though, that even if the IAT were susceptible to participants’ intentions to “beat the test,” it seems quite unlikely that this undermines the basic results. Implicit measures often reveal attitudes and stereotypes quite discrepant from self-reported ones. Finally, given that people want to minimize the appearance of bias—even implicit biases—if faking is going on, that just means that implicit biases in the population are in fact larger than reported.

⁸⁹ To demonstrate construct validity, researchers generally point out the ways in which some instrument shows both *convergent* and *discriminant* validities. As applied to the IAT, that would mean that the IAT scores converge with measures that we would expect it to converge with (e.g., other measures of attitudes and stereotypes), and depart from measures that we would expect it to depart from.

Implicit bias as measured by reaction-time techniques converges with physiological measures. For example, people with higher levels of implicit race bias showed more activation in the amygdala—an area of the brain associated with emotional, particularly fear, responses—when viewing unfamiliar Black (versus White) faces. See William A. Cunningham et al., *Neural components of social evaluation*, 85 J. Personality & Soc. Psychol. 639 (2003); Elizabeth A. Phelps et al., *Performance on Indirect Measures of*

the IAT measures not individual bias but salience asymmetry⁹⁰ or general cultural associations unattributable to any specific person⁹¹ have also been answered.

[50] This brief summary could not possibly detail the reasons and evidence in favor of reliability and validity in any comprehensive manner. Curious readers should turn to cited sources. Our point is not that the IAT is somehow immaculate. No psychological measure is perfectly pure; even a simple self-reported self-esteem scale consisting of responses to a statement such as “On the whole, I

Race Evaluation Predicts Amygdala Activation, 12 J. COGNITIVE NEUROSCIENCE 729, 729 (2000). These correlations were stronger when faces were presented subliminally (that is, they appeared on the screen for such a brief moment that viewers were not even aware they had been presented) than supraliminally (participants were aware of seeing the faces).

Additionally, Whites with greater levels of implicit racial bias showed more physiological stress when speaking to a Black audience than those with lower levels. See Wendy B. Mendes et al., *Why egalitarianism might be good for your health: Physiological thriving during inter-racial interactions*, 18 PSYCHOLOGICAL SCIENCE 991 (2007).

⁹⁰ See Klaus Rothermund & Dirk Wentura, *Underlying processes in the Implicit Association Test (IAT): Dissociating salience from associations*, 133 J. EXPERIMENTAL PSYCHOLOGY: GENERAL 139 (2004). The claim here is that particularly salient items are easier to group together. So, if Black is more salient than White (for example to White subjects), and if Bad is more salient than Good, then the Black-Bad association may be driven by salience, not a negative attitude toward Blacks.

For a responsive comment, see Anthony Greenwald et al. *Validity of the salience asymmetry interpretation of the IAT: Comment on Rothermund and Wentura* (2004), 134 J. EXPERIMENTAL PSYCHOLOGY: GENERAL 420, 420 (2005) (pointing out that salience asymmetries “have the potential to contribute to IAT effects, much as do any other features that afford a basis for distinguishing among categories” but that they cannot account for findings of predictive validity or the emergence of preferences of novel groups). For a reply to the comment, see Klaus Rothermund et al., *Validity of the salience asymmetry account of the Implicit Association Test: Reply to Greenwald, Nosek, Banaji, and Klauer* (2005), 134 J. EXPERIMENTAL PSYCHOLOGY: GENERAL 426 (2005).

⁹¹ See, e.g., Hal R. Arkes & Philip E. Tetlock, *Attributions of Implicit Prejudice, or “Would Jesse Jackson ‘Fail’ the Implicit Association Test?”*, 15 PSYCH. INQUIRY 257, 268-74 (2004); Hal R. Arkes & Phillip E. Tetlock, *The implicit prejudice exchange: Islands of consensus in a sea of controversy: Response*, 15 PSYCHOLOGICAL INQUIRY 311-321 (2004); Michael A. Olson & Russell H. Fazio, *Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT*, 86 J. PERSONALITY & SOCIAL PSYCHOLOGY 653 (2004); Andrew Karpinski & James Hilton, *Attitudes and the Implicit Association Test*, 81 J. PERSONALITY & SOCIAL PSYCHOLOGY 774 (2001).

For responses and analysis, see Eric Luis Uhlmann et al., *Automatic Associations: Personal Attitudes or Cultural Knowledge?*, in IDEOLOGY, PSYCHOLOGY, AND LAW (Jon Hanson, ed. 2010); Brian A. Nosek & Jeffrey J. Hansen, *The associations in our heads belong to us: Searching for attitudes and knowledge in implicit cognition*, 22 COGNITION AND EMOTION (2008); Mahzarin R. Banaji et al., *No place for nostalgia in science: A response to Arkes and Tetlock*, 15 PSYCHOLOGICAL INQUIRY 279 (2004).

We resist overly stylized demarcations that separate culture from person. Quite obviously, culture plays a huge role in feeding the associations in our brains. But if IAT scores simply reflected a tally of the associations seen in the world around us but revealed nothing about the individual, we would not expect them to systematically correlate with individual behavior, as they do. Moreover, it’s not clear how and why some of these cultural explanations should have moral or legal significance. See Jerry Kang, *Comment on Uhlmann et al., Automatic Associations*, in IDEOLOGY, PSYCHOLOGY, AND LAW (Jon Hanson, ed. 2010) (manuscript on file with authors).

am satisfied with myself”⁹² can be affected by myriad variables such as participants’ literacy, the comfort of the room in which they respond, and a success or failure they had immediately prior to replying. We simply suggest that the IAT is no different, and the evidence suggests that none of the potential confounds can account for IAT effects in the hundreds of studies that have used the task.

[51] To provide still further comfort that we are not pulling a fast one, we point out that if the IAT had no reliability and validity whatsoever, it should have absolutely no predictive power of real world behaviors. After all, if it’s a junk instrument—something like a “mind reading” oscilloscope from the 30th century made from tinfoil, sold on e-bay—then it shouldn’t correlate with anything. But as detailed below, the IAT has demonstrated predictive validity. In short, there is a there there, and although finer scientific questions can and should be debated, we believe outright “denial” is no longer credible.

2. Definitional Denial

[52] What’s really driving “denial” might not be a scientific as much as a definitional dispute. A skeptic could simply be arguing that the mental construct of “bias”—by definition—includes only those attitudes and stereotypes that we are consciously aware of *and* endorse.⁹³ By definitional fiat, “implicit bias” would become oxymoronic, like “jumbo shrimp.”

[53] But there is no good scientific reason for this *ipse dixit*. In psychology, other mental constructs such as attention, perception, and memory all have explicit and implicit counterparts. For instance, we do not reject the concept of implicit memory simply because the memory cannot be consciously recalled. Why, then, carve out attitudes and stereotypes for disparate treatment?⁹⁴ From a linguistic perspective, we do no violence to the English language in using the words “attitude” and “stereotype” in labeling implicit social cognitions. The same goes for calling departures from a neutral point “bias.”⁹⁵ These are standard scientific deployments of the terms.⁹⁶

⁹² This is an item from the Rosenberg self-esteem scale. See MORRIS ROSENBERG, *SOCIETY AND THE ADOLESCENT SELF-IMAGE* (1965).

⁹³ Hal Arkes and Philip Tetlock appear to do this when they argue that a mere mental association should not be portrayed as an automatic attitude because doing so “converts an association one has to an attitude *one endorses at some level*.” Arkes & Tetlock, *supra* note 91, at 268 (emphasis added).

⁹⁴ See Nosek et al., *Nostalgia*, *supra* note 91, at 280.

⁹⁵ Hart Blanton and James Jaccard have questioned whether a score of “zero” on the IAT truly reflects no bias, using data simulations to suggest that the IAT is on an arbitrary scale. They argue that “[f]eatures of the IAT make it unusually bold to assume that the observed measure score of 0 reflects the true rational zero-point.” Harton Blanton & James Jaccard, *Tests of multiplicative models in psychology: A case study using the unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept*, 113 *PSYCHOLOGICAL REVIEW* 155, 157 (2006). In subsequent data reanalysis and simulations, Greenwald and colleagues concluded that the problems brought up by Blanton and Jaccard were “much more delimited than implied . . . [and a] plausible complicating factor in only one of the five experiments analyzed by Greenwald et al.” Anthony G. Greenwald, *Why so little faith? A reply to*

[54] Of course, we have already observed that having implicit cognitions about broccoli does not threaten our flattering self-portraits, whereas implicit cognitions about, say, the competence of Asian women might.⁹⁷ But unpleasantness rarely says anything about scientific truth.

B. Minimization

[55] Outright denial is no longer tenable because the data are too numerous and findings have become mainstream. The next predictable phase, then, is to accept the existence of implicit social cognitions but to argue their unimportance. Stated strongly, implicit social cognitions do not influence real world behavior whatsoever,⁹⁸ and are thus of only academic interest.⁹⁹ Stated more weakly, implicit biases influence behavior marginally, but are easy to correct for, and by the way, the status quo is already doing it. This mode of response is called *minimization*.

1. No real-world behaviors

[56] Having discussed how implicit bias could have influenced *Juan's Interview*, *Skip's Encounter*, and *Grace's Hire*, the strong form of minimization seems farfetched. We hammer this point deeper by exploring Juan's storyline.

[57] Before Juan even got the interview, he had to submit a resume. Back in 2004, behavioral economists Marianne Bertrand and Sendhil Mullainathan produced an important audit study. They sent comparable resumes out to numerous employers in Boston and Chicago, but used names such as "Emily" or "Greg" to signal Whiteness and "Lakisha" and "Jamal" to signal Blackness.¹⁰⁰

Blanton and Jaccard's (2006) skeptical view of testing pure multiplicative theories, 113 *PSYCHOLOGICAL REVIEW* 170, 178 (2006). See also *supra* note 79.

⁹⁶ See Anthony Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 *CALIF. L. REV.* 945, 952-53 (2006) (discussing various definitions and usages of "bias").

⁹⁷ See *supra* ¶ [17].

⁹⁸ By "behavior," we adopt the standard psychological distinction between mental constructs, on the one hand, such as attitudes and stereotypes, and some behavioral manifestation, on the other hand, which includes differential evaluations, judgments, and physical behaviors.

⁹⁹ It's, of course, okay for academics to be interested in matters of only academic interest. Cf. Mahzarin R. Banaji & Robert G. Crowder, *The Bankruptcy of Everyday Memory*, 44 *AMERICAN PSYCHOLOGIST* 1185, 1191 (1989) (explaining in the context of memory research that the "impatience to know what practical application justifies every popular laboratory protocol" might stem from a failure to distinguish "between science and technology").

¹⁰⁰ Marianne Bertrand & Sendhil Mullainathan, *Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination*, 94 *AMER. ECON. REV.* 991 (2004).

They found that the trivial manipulation of name produced a 50% difference in callback rates.¹⁰¹

[58] In 2007, Dan-Olof Rooth replicated and extended this experiment by connecting the resume sort to implicit bias. He sent 1,152 resumes to posted job listings in Sweden; resumes were again created to be similar in background, experience, and qualifications but to differ in the applicant's ethnicity.¹⁰² By altering the last name, each resume either denoted an Arab/Muslim job candidate (whose experience, education and short biography made clear that he was born and educated in Sweden) or a Swedish job applicant.

[59] In 283 cases, only one of the applicants was invited for an interview. And in these cases, a clear preference emerged for the candidate with the Swedish-sounding name (217 invites) over the one with the Arab/ Muslim-sounding name (66 invites). In other words, the identically qualified candidate was 3.3 times more likely to be called back simply because he enjoyed a Swedish name. If we consider all the cases (including those in which none or both were invited), candidates with Swedish-sounding names had an approximately 9% advantage in callback rates.¹⁰³

[60] Several months later, Rooth located 193 of the recruiters or managers responsible for the hiring decision. They reported explicit attitudes and stereotypes and their hiring preferences,¹⁰⁴ and completed an IAT assessing the Arab-Muslim male stereotypes. Recruiters generally reported some explicit preference for both hiring Swedish men (54% of recruiters) and having warmer feelings toward them (45% of recruiters). Although they did not, on average, explicitly endorse the stereotype that Swedish men were more productive than Arab-Muslim men, the IAT revealed a strong implicit association between Swedish men and "high productivity."¹⁰⁵

[61] The critical empirical question was the extent to which implicit stereotypes predicted the callback disparity. Among the firms that had invited at least one candidate to interview, higher levels of implicit bias predicted discrimination. A one standard deviation increase in implicit bias translated into an approximately 12% decrease in the probability that an Arab-Muslim candidate received an interview.¹⁰⁶ Put another way, approximately half of the large discrepancy

¹⁰¹ See *id.* at 998.

¹⁰² See Dan-Olof Rooth, *Implicit Discrimination in Hiring: Real World Evidence* (Institute for the Study of Labor, Discussion Paper No. 2764 April 2007).

¹⁰³ See *id.* at 5. In 239 cases, both applicants were invited to interview.

¹⁰⁴ They indicated strong, moderate, or some preference for hiring "Arab-Muslim men (in Sweden)" and native Swedish men. *Id.* at 6.

¹⁰⁵ See *id.* at 8-9.

¹⁰⁶ Among all firms, including those that had extended no interview offers, a one standard deviation increase was associated with a smaller (3%) but still statistically significant decrease in the probability that an Arab-Muslim candidate would receive an interview. See *id.* at 11-12. In both cases, the explicit

between interview invitations for men with Arab-Muslim names and men with native Swedish names could be accounted for by implicit bias, and approximately one quarter could be accounted for by explicit bias.¹⁰⁷

[62] Let's tie this back to Juan's story. His name alone might make it harder to get an interview perhaps because of implicit stereotypes regarding work ethic. If Juan is lucky enough to be called in, then the social interaction may sour, in part because of negative implicit attitudes. And if he were to get the job, substantial research suggests that managers will interpret ambiguous performance in ways that confirm prior expectations.¹⁰⁸ We do not want to overstate the case. If Juan is truly stellar, he will likely be celebrated; if Juan is truly incompetent, again, subtle biases are the least of his worries. But if he is somewhere in the vast undifferentiated middle—like most of us are—then implicit biases may impact his work trajectory.

[63] Instead of running through a tiresome catalog of still more predictive validity studies, we instead rely on two summaries. In a recent paper aimed at organizational managers, John Jost and colleagues¹⁰⁹ respond to critiques of the implicit bias science by Philip Tetlock and Gregory Mitchell.¹¹⁰ Jost et al. point out that summary dismissal of the research on implicit bias would be tantamount to disputing the “major tenets of 20th century cognitive psychology.”¹¹¹ They review “10 studies that no manager should ignore” showing the effects of implicit bias in studies like the ones reviewed.

[64] If ten studies aren't enough, how about an order of magnitude more? Anthony Greenwald and colleagues have completed just such a meta-analysis of 122 research reports, including 184 independent samples, involving 14,900 subjects, that included the implicit association test (IAT) and participant behaviors.¹¹² A meta-analysis stitches together the weighted findings of all

measures did not significantly predict decisions, although Rooth noted that such bias may nevertheless be “economically important.” *Id.* at 13.

¹⁰⁷ See *id.* at 16

¹⁰⁸ See, e.g., John M. Darley & Paget H. Gross, *A Hypothesis-Confirming Bias In Labeling Effects*, 44 J. PERSONALITY & SOCIAL PSYCHOLOGY 20 (1983). In this study, cues about a child's social class affected interpretation of her academic performance. People who believed she was from a higher socioeconomic status described her performance as strong, whereas those who believed she was from a lower socioeconomic status described the identical performance as weaker. *Id.* at 27.

¹⁰⁹ Jost et al., *The existence of implicit is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore*, RESEARCH IN ORGANIZATIONAL BEHAVIOR (B.M. Staw & A.P. Brief, Eds. in press) (manuscript on file with authors).

¹¹⁰ See Tetlock & Mitchell (same volume)

¹¹¹ Jost, *supra* note 109, at 13.

¹¹² 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) (hereinafter *Meta-analysis*).

studies within a designated subject matter and timeframe, to produce conclusions based on the broadest possible evidentiary foundation. Any single study will have some methodological imperfection. But a meta-analytic approach will not be confounded by such faults unless they appear systematically throughout the dataset, which is far less likely.

[65] This meta-analysis revealed that both explicit biases (measured by self-reports on surveys) and implicit biases (measured by the IAT) predicted various criterion variables, such as nonverbal behavior, social judgments, physiological responses, and social action. The correlations for both types of bias were moderate in size, with explicit biases performing slightly better in general prediction.¹¹³ However, in socially sensitive domains of behavior, the IAT predicted criterion behavior *better* than explicit self-reports did.¹¹⁴

[66] How good of a job did it do? In the White-Black discrimination domain (i.e. whether anti-Black implicit attitude predicts behavior toward Blacks), implicit bias predicts 5.7% of the variability in behavior.¹¹⁵ Now, one might retort that this number is too low to have policy significance.¹¹⁶ But the correlation between implicit bias and behavior may be substantially higher, and masked by imperfections in the instruments that measure both the bias and the behavior.¹¹⁷

¹¹³ The correlation for implicit bias was average $r=0.274$; the correlation for explicit bias was a slightly larger average $r=0.361$.

¹¹⁴ This was the case for White-Black discrimination, where the implicit bias correlations were average $r=0.24$, significantly higher than explicit bias correlations of average $r=0.12$. It was also the case for other intergroup discriminations (dealing with other ethnic groups, age, weight), in which implicit biases correlated with behavior on an average $r=0.20$, whereas explicit biases correlated on average $r=0.12$. See *id.* at 73 table 3.

¹¹⁵ The correlation is $r=0.24$. By squaring the r ($r^2=5.7\%$), we get the percentage of variance explained. It reveals the extent to which differences in behavior can be attributed to a single variable (implicit bias).

¹¹⁶ See Gregory Mitchell & Philip Tetlock, *Facts Do Matter: A Reply to Bagenstos*, 2009 Va. L. School Public L. & Leg. Theory Working Paper Series #114, p.23 (2009) ("If we accepted this estimate as reliable, the psychometric fact that IAT scores have low positive correlations with behavior expansively defined as discrimination would guarantee that many individuals labeled implicitly biased by the IAT will not exhibit discriminatory behavior of any kind.")

¹¹⁷ If we assume that implicit measures have a test-retest reliability of approximately .56, and that the behavior has a reliability of .80, the correlation between implicit bias and behavior generally (not just in socially sensitive domains) could rise from 0.274 to what is called a *disattenuated correlation* of 0.41. This would then explain 16.7% of the variability in behavior. To focus again on just behaviors within socially sensitive domains such as Black-White discrimination, a disattenuated correlation could rise from 0.24 to 0.36.

Brian A. Nosek et al., *The Implicit Association test at Age 7: A methodological and Conceptual Review*, in *AUTOMATIC PROCESSES IN SOCIAL THINKING AND BEHAVIOR* 265 (JOHN A. BARGH, ED. 2007), 274. (providing an estimate for our assumption of the reliability of implicit measures whereas the assumption of the reliability of behavior is an estimate with no strong empirical basis (but is the same number used by Greenwald et al., *supra* note 112, at 28, in their similar calculation). The formula for this new variable—the correlation corrected for attenuation = $r_{xy}/\sqrt{(\text{Measure } x_{\text{reliability}} \times y_{\text{reliability}})}$. Notably,

[67] Even if this is not the case, explaining 5.7% of the variability in Black-White discrimination should be seen as meaningful. Here's a baseball analogy. With two outs in the ninth winning, with a man on second base, who do you want at the plate: All-Star Matt Holliday (.321 average in 2008) or John Buck (.224 average in 2008)? But in statistical terms, Robert Abelson has demonstrated that the batting average difference explains only 1.3% of the variance in a single at-bat (hit or no-hit).¹¹⁸ This counterintuitive result stems in part from the fact that we are looking at one player and one at-bat, instead of a string of players batting in a game, over an entire season.¹¹⁹ If we extend the baseball analogy back to a social category such as race, then we should remember that we are summing up all the differential treatment over all interactions and over all time over all the people within each racial "team." From this vantage point, even low-level correlations can generate significant consequences, especially as measured by average group differences.

2. Easy correction

[68] The weaker version of minimization suggests that in the real world, virtuous individuals can simply will themselves to be fair and not allow their behavior to be altered by implicit biases. In addition, organizational practices, which include diversity training, regularized procedures, supervision, and accountability prevent behavioral manifestation of implicit biases.¹²⁰ In short, the status quo outside of the laboratory is already correcting any problem and doing so easily.

[69] On the one hand, implicit biases are malleable and the environment can strongly influence how and whether implicit biases translate into behavior.¹²¹

as the reliability of the measures decreases, the corrected correlation increases. Thus, an estimate of behavior's reliability of .80 is a conservative choice in calculating the disattenuated correlation.

¹¹⁸ See Robert P. Abelson, *A variance explanation paradox: When a little is a lot*, 97 PSYCHOLOGICAL BULLETIN 129, 132 (1985). Abelson used a simulation of a .320 hitter (an exceptional performance achieved by only eight players who had at least 375 plate appearances during the course of the 2008 major league baseball season) and a .220 hitter (a mediocre performance achieved by 221 of the 226 players with at least 375 plate appearances during the 2008 season).

¹¹⁹ As Abelson cautioned:

[T]he attitude toward explained variance ought to be conditional on the degree to which the effects of the explanatory factor cumulate in practice In such cases, it is quite possible that small variance contributions of independent variables in single-shot studies grossly understate the variance contribution in the long run.

Id. at 133.

¹²⁰ See, e.g., Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 OHIO ST. L.J. 1023, 1109-10 (2006) (listing twelve factors why laboratory findings might not generalize to the real world);

¹²¹ For a general discussion of the debiasing literature, see, e.g., Adam Benforado & Jon Hanson, *The Great Attributional Divide: How Divergent Views of Human Behavior Are Shaping Legal Policy*, 57 EMORY L.J. 311, 334-37 (2008).

On the other hand, it seems unduly optimistic to think that this is easy to do¹²² and is already taking place.

[70] Being motivated to be egalitarian should influence whether implicit bias translates into discriminatory behavior.¹²³ Indeed, data confirm this. Implicit bias as measured by priming techniques predicted differential trait ratings of Blacks (relative to Whites) for those with little motivation to control bias; however, this was *not* the case with those highly motivated to control bias.¹²⁴ Similar results have been found with IAT-measured implicit bias.¹²⁵ The problem is that we won't be motivated to check bias if we are cocksure that we are bias-free. Unfortunately, we readily see bias in others but not in ourselves.¹²⁶

[71] This analysis cross-applies to institutions. To be sure, the accountability literature reveals that individuals who have to explain their decisionmaking to others are less prone to various biases. And, for example in the workplace, today's human resources departments and supervisors do have to account to their superiors for employment actions. That said, organizations are not asking their decisionmakers to account for implicit-bias actuated discrimination. And given the illusion of objectivity¹²⁷ and the difficulty of spotting in any isolated case whether implicit bias was a but-for (much less proximate) cause for some decision, it seems heroic to assume that general accountability pressures are automagically solving the problem.

¹²² See John A. Bargh, *The Cognitive Monster: The Case Against the Controllability of Automatic Stereotype Effects*, in DUAL-PROCESS THEORIES IN SOCIAL PSYCHOLOGY 361 (SHELLY CHAIKEN & YAACOV TROPE, EDS. 1999) (hereinafter "DUAL-PROCESS THEORIES").

¹²³ See Russell H. Fazio & Tamara Towles-Schwen, *The MODE Model of Attitude-Behavior Processes*, in DUAL-PROCESS THEORIES, *supra* note 122, at 97.

¹²⁴ In fact, the latter category showed the opposite pattern, suggesting the possibility of overcorrection. See Michael A. Olson & Russell H. Fazio, *Trait inferences as a function of automatically-activated racial attitudes and motivation to control prejudiced reaction*, 26 BASIC AND APPLIED SOCIAL PSYCHOLOGY 1 (2004); Tamara Towles-Schwen & Russell H. Fazio, *Choosing social situations: The relation between automatically- activated racial attitudes and anticipated comfort interacting with African Americans*, 29 PERSONALITY AND SOCIAL PSYCHOLOGY BULLETIN 170 (2003). Similarly, for White participants low in motivation to control prejudicial responses, implicit bias predicted anticipated comfort level during an unscripted interaction with a Black partner. However, it was not predictive for those strongly motivated to control prejudice. See *id.* at 176-178.

¹²⁵ For example, implicit bias toward gay people predicted nonverbal behavior during an interaction with a gay partner only for participants low in motivation and in the tendency to control their behavior. See Nilanjana Dasgupta & Luis M. Rivera, *From automatic antigay prejudice to behavior: The moderating role of conscious beliefs about gender and behavioral control*, 91 J. PERSONALITY & SOC. PSYCHOL. 268 (2006).

¹²⁶ Emily Pronin et al., *The bias blind spot: Perceptions of bias in self versus others*, 28 PERSONALITY AND SOCIAL PSYCHOLOGY BULLETIN 369 (2002). For a review see Emily Pronin, 320 SCIENCE 1177 (2008).

¹²⁷ See generally David Alain Armor, *The Illusion of Objectivity: A Bias in the Perception of Freedom From Bias* (1998) (unpublished dissertation); Nandita Murukutla & David A. Armor, *Illusions of Objectivity and the Dispute over Kashmir. An Experimental Test of the Effects of Disagreement* (Oct. 7, 2004) (unpublished manuscript). See also

[72] Indeed, the evidence is to the contrary. Recalling *Juan's Interview*, if accountability to antidiscrimination laws and policies were solving the problem, Jamal should have been called back as often as Greg. Returning to *Skip's Encounter*, surely our men and women in blue feel all kinds of accountability not to shoot an innocent person. But in at least one set of “shooter bias” studies, police officers responded like civilians and were more likely to shoot unarmed Black than unarmed White targets.¹²⁸

C. Politicization

[73] As evidence of real world impact amasses, the stakes will grow larger. It is inevitable that various parties will then embrace or attack the science to further particular agendas. The general name for this mode of engagement is “politicization.” These reactions will come from the political Left and Right.

1. Overreaching by the Left

[74] The implicit bias findings dispute any notion that we are literally colorblind or gender-blind. Emboldened by such research, some on the Left might offer conclusions that extend too far beyond the data.¹²⁹

[75] First, the findings from implicit social cognition may be mischaracterized as equivalent—both morally and psychologically—to traditional Bull Connor-like¹³⁰ animus. But implicit social cognition is not within one’s consciousness, and would often be repudiated if considered. By contrast, Bull Connor was well aware of his explicit preferences and embraced them. Accordingly, we caution

¹²⁸ E. Ashby Plant & B. Michelle Peruche. *The consequences of race for police officers' responses to criminal subjects*. 16 *Psychological Science* 180, 181 (2005). In another paper, police officers showed a similar tendency to be faster to respond to armed Blacks (compared to armed Whites) and unarmed Whites (compared to unarmed Blacks), but did not exhibit racial bias on the arguably more important criterion of accuracy: unarmed Blacks were not more likely to be “shot” than unarmed Whites. See Joshua Correll et al. *Across the thin blue line: Police officers and racial bias in the decision to shoot.*, 92 *J. PERSON. & SOC. PSYCH.* 1006, 1010-1013, 1016-1017 (2007) (describing results from two studies).

¹²⁹ In this discussion, we are not accusing specific individuals of such mistakes. That is partly because such mistakes are not regularly made, certainly not by professional scholars. That said, some such overreaching should not surprise us—especially in blogs and other popular media. Just because some on the Left might make such mistakes does not mean that everyone on the Left should be smeared with “guilt by association.”

¹³⁰ Theophilus Eugene “Bull” Connor was a Commissioner of Public Safety in Birmingham, Alabama. His biases were not implicit. He used dogs and fire hoses to attack civil rights marchers. See WILLIAM A. NUNNELLY, *BULL CONNOR* 163-64 (1991).

against such slippage.¹³¹ This explains why implicit social cognition scholars have avoided inflammatory labels such as implicit “bigots” or “racists.”¹³²

[76] Second, another way to overreach is to focus solely on the implicit biases held by members of socially privileged groups, such as straight, young, wealthy, Christian, light-skinned, White, able-bodied males. But this ignores the pervasiveness of implicit bias, including its internalization by members of subordinate groups.¹³³ We also highlight, as in Figure 1, evidence of substantial implicit bias from specific minority groups against other minority groups.¹³⁴ Being at the bottom of some relative social hierarchy does not immunize us from implicit bias, against ourselves or others in similar subordinated positions.¹³⁵

[Insert Figure 1 about here]

[77] Third, overenthusiasts might characterize the IAT as a “racism/sexism polygraph” that detects prejudice or stereotypes, which should be used to measure specific individuals for hiring or selection. To be sure, employers regularly use selection tools that may have weaker scientific foundations than

¹³¹ See, e.g., Kang & Banaji, *Fair Measures*, *supra* note 2, 1077 (“We are not arguing that discrimination caused by implicit bias should be *equally* problematic as that caused by consciously endorsed explicit bias.”); *id.* (“We are not arguing that implicit bias-induced discrimination should produce the same legal liability as explicit animus-driven discrimination under current equal protection doctrine or federal antidiscrimination statutes....”); Nosek et al., *Nostalgia*, *supra* note 91, at 281.

¹³² See, e.g., Frequently Asked Question (FAQ) 16 on Project Implicit:

16. If my Black-White attitude IAT shows automatic White preference, does that mean that I’m prejudiced?

Answer: This is a very important question. Social psychologists use the word ‘prejudiced’ to describe people who endorse or approve of negative attitudes and discriminatory behavior toward various out-groups. Many people who show automatic White preference on the Black-White attitude IAT are not prejudiced by this definition. <https://implicit.harvard.edu/implicit/demo/background/faqs.html> (visited 2007 10 21) (emphasis added).

¹³³ Although African Americans associate African American (versus White American) with weapons (versus harmless object) at a lower level than other racial groups, they still have a fairly strong association. Nosek et al., *supra* note 42, at 20 ($d=0.59$). Asian Americans associate Asian faces (versus White faces) with foreign (versus American) at a lower level than other racial groups, but they too show that implicit association—even when they are told that all the Asian faces they see in an IAT are of American citizens and all the European faces they see are of European citizens. *Id.* at xx ($d=0.29$). In the domains of gender, i.e. whether women or men are more associated with science (versus liberal arts) or with career (versus family), women show larger implicit biases than men. *Id.* at 12 (gender science women $d=0.98$, men = 0.93; gender career women $d=1.19$, men = 0.94).

¹³⁴ Both Asians and Latinos, for instance, show substantial attitudinal preference for Whites over Blacks. Indeed, the magnitude of White attitudinal preference follows the racial hierarchy in American today: Whites shows the greatest pro-White (as compared to Black) attitudes, then Asians, Latinos, American Indians, and Blacks.

¹³⁵ See Nosek et al., *supra* note 42. As seen in Figure 2, Black subjects’ show an average tendency to slightly prefer Black over White. Approximately half of all Black test-takers showed an implicit preference for White over Black.

the IAT, such as the Myers-Briggs personality test,¹³⁶ free form interviews, “chemistry,” or other “gut” feelings. And there has been no systematic outcry to eradicate those tests as scientific abominations. That said, on the state of the current evidence, the leading implicit bias scholars have recommended *against* using the IAT in individual selection context.¹³⁷

[78] Fourth, Leftists might overclaim the reformist implications of the research. For example, they might claim that the demonstration of these biases and their behavioral manifestations show that all affirmative action programs across the board should be preserved. But that conclusion does not derive from the evidence. There may be various “fair measures”¹³⁸ worth taking now to counter implicit bias; however, various other affirmative action programs, for example those sounding in corrective justice, cannot be justified on these particular scientific grounds. Other powerful justifications exist, but they aren’t based directly on implicit social cognition.¹³⁹

2. Backlash from the Right

[79] Backlash looks substantively like the “denial” and “minimization” modes already discussed. But in style, Right-wing backlash is more aggressive and ad hominem. It characterizes the implicit bias research as junk science,¹⁴⁰ politi-

¹³⁶ The use of Myers-Briggs as a tool for career advising and sometimes even personnel selection persists despite debate among academic research psychologists to its validity or usefulness. See, e.g., John Hunsley et al., *Controversial and questionable assessment techniques*, in SCIENCE AND PSEUDOSCIENCE IN CLINICAL PSYCHOLOGY (SO LILIENFELD ET AL., EDS. 1986); Bruce Thompson & Gloria M. Borrello, *Construct Validity of the Myers-Briggs Type Indicator*, 46 EDUC. & PSYCHOL. MEASUREMENT 745, 750-51 (1986) (finding construct validity on the basis of factor analysis).

¹³⁷ Recently, Mahzarin Banaji testified under oath in the New Hampshire Supreme Court against using the IAT in individual selection contexts, such as selecting the jury. See also *supra* ¶ [48] (discussing reliability).

¹³⁸ See Kang & Banaji, *supra* note 2 (outlining concrete suggestions).

¹³⁹ To be fair and balanced, overreaching can also come from the Right. For example, one could imagine a White supremacist celebrating the pervasiveness of implicit biases. They could use the data as evidence that the propaganda of a politically correct Nanny state has failed to eradicate fundamental, indeed genetically-determined truths. Indeed, in the rare instances in which a neo-Nazi respondent has communicated with Project Implicit researchers, they have done so to share their gratification that the test “proved” that they implicitly held their explicitly endorsed biases.

¹⁴⁰ For example, Gregory Mitchell and Philip Tetlock suggest that the work of the “implicit prejudice scholars” (we prefer the more accurate and less-loaded term “implicit bias” or “implicit social cognition”) shouldn’t be called “science.” See Mitchell & Tetlock, *supra* note 120, at 1029 (“Accordingly, implicit prejudice scholars work hard to claim the *mantle of science* as they advance their agenda.”) (emphasis added); *id.* at 1031 (noting that “some . . . claim a presumption of correctness in their interpretations of this ambiguous evidence by *attaching the label ‘science’* to their views.”) (emphasis added); *id.* at 1029 (describing scientific rhetoric as “more honorific than descriptive”).

At times, they hint that the entire field of psychology lacks credibility since it’s subject to fashionable (not evidence-based) whims. See *id.* at 1028 n.17 (describing the work of Charles Lawrence as “*unfashionable* among psychological theorists”) (emphasis added); *id.* at 1041 (“the focus has shifted with prevailing intellectual *fashions* from psychodynamic theories to social-identity theories”); *id.* at 1062 (“by *now-out-of-fashion* psychodynamic theories”).

cally motivated by junk scientists and their lawyer accomplices, who manipulate data, misinterpret results, and exaggerate findings¹⁴¹ in order to snooker society into politically correct¹⁴² wealth transfers. By contrast, its own call for skepticism is described as disinterested, high-minded “sound science” that simply asks for rigorous and faithful adherence to scientific orthodoxy.

[80] What’s ironic is that this embrace of science may be incidental: the primary objective is to block change. The tactic, however, is to demand scientific rigor that is unachievable while conveniently forgetting to require the same of factual assumptions undergirding the status quo. Consider, for example, the tobacco industry’s response to cancer findings. According to Chris Mooney, as soon as the Environmental Protection Agency (EPA) concluded that secondhand smoke killed,¹⁴³ the tobacco industry generated its own counter-science contesting the causation evidence. It complained that the causation science was junk, unproven, insufficiently rigorous. In short, it manufactured scientific “doubt,” a deliberate strategy deployed by the tobacco industry for decades.¹⁴⁴ The tobacco industry even impugned the EPA’s scientific integrity, as “another step in a long process characterized by a preference for political correctness over sound science.”¹⁴⁵ The irony, of course, was that the tobacco executives did not much care about the substantive scientific merits; rather, their principal objective was to sell smokes.

[81] We want not to be misread: We are *not* suggesting that any call for skepticism that challenges our particular scientific understanding, based on the best evidence as of 2009, should be tagged “backlash.” To the contrary, skepticism is central to the practicing scientist’s attitude toward problem solving and discovery. After all, it was skepticism about explicit self-reports that

Mitchell has criticized psychological approaches in other contexts, in defense of Rational Choice Theory. See, e.g., Gregory Mitchell, *Why Law and Economics' Perfect Rationality Should Not Be Traded for Behavioral Law and Economics' Equal Incompetence*, 91 GEO. L. J. 67, 131 (2002) (“As currently conceptualized, legal decision theory is nothing more than a mess of overgeneralizations about how people exhibit this or that bias or anomaly under largely unspecified conditions.”).

¹⁴¹ Mitchell and Tetlock lament the “superficial and selective” reference to relevant factors, *id.* at 1091; refusal to provide “full disclosure”, *id.* at 1092; and deploying “specious but seductive” arguments, *id.* at 1100, that they reveal by “pulling back the curtain,” *id.* at 1030. Criticizing those who use behavioral economics, Mitchell has suggested that this sort of exaggeration is done because it “probably sells better than a nuanced, contextualized picture of human behavior . . . lacking in cognitive universals.” Gregory Mitchell, *Taking Behavioralism Too Seriously? The Unwarranted Pessimism of the New Behavioral Analysis of Law*, 43 Wm. & Mary L. Rev. 1907, 2018.

¹⁴² Mitchell and Tetlock complain that implicit bias scholars are driven by “moral certitude”, Mitchell & Tetlock, *supra* note 120, at 1029, come from an “ideologically-skewed field,” *id.* at 1064, engaged in “politicized research programs in which hypothesis advocacy has supplanted hypothesis testing,” *id.* at 1076, without “political even-handedness” trying to “co-opt a value-laden concept to advance [our] policy agenda”, while “repeated[ly] fail[ing] . . . to acknowledge the role that political values unavoidably play” in the analysis.

¹⁴³ See CHRIS MOONEY, *THE REPUBLICAN WAR ON SCIENCE* 66 (2005).

¹⁴⁴ *Id.* at 67.

¹⁴⁵ *Id.* at 66.

drove the discovery of implicit social cognition. Also, methodological concerns and critiques spurred real advances in the methodology and scoring procedures used with the IAT.¹⁴⁶ The earlier discussion of denial and minimization were substantive engagements in precisely this vein—both respectful and inviting on-the-merits skepticism. That said, backlash scholarship is different.

[82] What we need is a way to distinguish between legitimate (even if heated) disagreements and backlash, even when self-interested reasoning abounds.¹⁴⁷ Although there is no surefire way, we have some suggestions. First, questions of scientific correctness should be decoupled from allegations of venality or fraud. Second, we should respect the relative institutional competencies of the scientific versus legal communities. Journals that are peer-reviewed by experts are the proper fora for the correctness of scientific questions. We see no reason why the operation of scientific orthodoxy, over the long run, would not produce a consensus. It did so with global warming. By contrast, general law reviews are suitable fora to resolve allegations of bad faith, intentional misuse of evidence, mischaracterization of research, reckless disregard for precision, and ideological advocacy. If mud is being thrown, lawyers know how to respond.¹⁴⁸ In addition, law reviews are a fine place to engage in normative debates, which are significant even when the descriptive account is agreed upon.

D. Resignation

[83] Even if implicit biases have real-world consequences, what if they could not be countered in any way? Perhaps these associations are “hardwired” into our brains. After all, if we have an implicit attitudinal preference for flowers over insects, perhaps it comes from hundreds of thousands of years of natural

¹⁴⁶ Critiques from academic psychologists spurred research on the cognitive mechanisms and limitations of the IAT, and hastened the development of superior methodological procedures, and scoring techniques. See Brian A. Nosek et al. *Understanding and using the Implicit Association Test II: Method variables and construct validity*. 21 PERSONALITY AND SOCIAL PSYCHOLOGY BULLETIN (2005); Anthony A., Greenwald et al. *Understanding and using the Implicit Association Test I: An improved scoring algorithm*. 85 J. OF PERSONALITY AND SOCIAL PSYCHOLOGY. (2003). See also *supra* note 87.

¹⁴⁷ For a discussion in the law reviews of self-serving reasoning, see Adam Benforado & Jon Hanson, *Naive Cynicism: Maintaining False Perceptions in Policy Debates*, 57 EMORY L.J. 499, 513-34 (2008).

¹⁴⁸ See, e.g., Samuel R. Bagenstos, *Implicit Bias, “Science,” and Antidiscrimination Law*, 1 HARV. L. & POL. REV. 477, 480 (2007) (arguing that the critique made by Gregory Mitchell and Phil Tetlock “is thus best understood, not as a scientific critique ... but as an argument about the normative bases for antidiscrimination law”); Adam Benforado & Jon Hanson, *Legal Academic Backlash: The Response of Legal Theorists to Situationist Insights*, 57 EMORY L.J. 1087, 1119 n.119, 1135 (2008) (describing Gregory Mitchell as engaging in backlash); Russell Korobkin, *Possibility And Plausibility In Law And Economics*, 32 Fla. St. U. L. Rev. 781, 782 n.5 (2005) (criticizing Gregory Mitchell’s strawman tendencies); Robert A. Prentice, *Chicago Man, K-T Man, and the Future of Behavioral Law and Economics*, 56 VAND. L. REV. 1663, 1744-45 (2003) (noting that “a large number of straw men were born and killed in the construction of [Gregory] Mitchell’s arguments”); Jeffrey J. Rachlinski, *Empirical Legal Realism: A New Social Scientific Assessment Of Law And Human Behavior: The Uncertain Psychological Case For Paternalism*, 97 NW. U. L. REV. 1165 (2003) (describing Gregory Mitchell’s strawman technique as “cheap academic stunt”).

selection. And, if these associations are immutable, we might as well resign ourselves to our biological limitations. Whether these arguments sound in sociobiology, evolutionary psychology, or behavioral genetics, this mode of response can be called “resignation.”

[84] The *resignation* mode features two components: (i) strict determinism (one could not have done otherwise), and (ii) some genetic or evolutionary origin story for that determinism. The second component—the origin story—grabs all the attention. But the first component—strict determinism—is what’s important. Most people believe that if one cannot do otherwise, we should be resigned to the inevitable. By contrast, if one can do otherwise, then we should decide *what* to do, largely independent of the origin story of implicit biases (regardless of any adaptive provenance).¹⁴⁹ To be sure, the origin story may help us reach our desired ends more efficaciously. But that is a “how” question more than a “what” question. In sum, we should concentrate on the first component: strict determinism.

[85] Having focused, let’s clarify the surrounding empirical picture. First, there is great individual variability in implicit biases.¹⁵⁰ Some people show strong bias, but others demonstrate minimal bias or even scores in the opposite direction of most test-takers. Of course, variability of implicit bias does not refute a claim of determinism. The point here is simply to contest crude claims about uniform biases predestinating our actions in some gross mechanical way.

[86] Second, even if implicit biases themselves cannot change, the causal link between biases and behavior can be disrupted through procedural and structural reforms. A stylized example comes from the hiring context. Up through the 1970s, female musicians were not being hired in major orchestras. After the adoption of “blind” auditions, the number of women hired jumped dramatically. In other hiring contexts, one could imagine “blinding” the names of all

¹⁴⁹ We do not profess to be even amateur philosophers, so our claims here are simply descriptive. We add that they seem to us to be reasonable moral stances.

¹⁵⁰ The standard deviations shown in Table 1 disclose both prevalence of implicit bias *and* substantial variance. Also, some people are more dispositionally motivated to be non-prejudiced, a tendency that moderates implicit social cognitions. People motivated to be non-prejudiced for personal (or internal) reasons, but not social (or external) reasons, showed reduced implicit racial bias on a physiological measure, see David M. Amodio et al., *Individual differences in the activation and control of affective race bias as assessed by startle eyeblink and self-report*, 84 J. PERSONALITY AND SOCIAL PSYCHOLOGY 738 (2003), and a reaction-time task, see Patricia G. Devine et al., *The regulation of explicit and implicit race bias: the role of motivations to respond without prejudice*, 82 J. PERSONALITY AND SOCIAL PSYCHOLOGY 835 (2002). For a case where motivation was related to a reaction-time, but not a physiological measure of bias, see Eric J. Vanman, *Racial discrimination by low-prejudiced Whites*, 15 PSYCHOLOGICAL SCIENCE 711 (2004). Implicit bias is also related to more general cognitive styles, such that people with highly rigid thinking styles or strongly right-wing ideologies exhibit stronger implicit bias (See William A. Cunningham et al. *Implicit and explicit ethnocentrism: revisiting the ideologies of prejudice*, 30 PERSONALITY AND SOCIAL PSYCHOLOGY BULLETIN 1332 (2004)).

resumes in initial sorts.¹⁵¹ One cannot even implicitly favor Joshua over Juan if neither name appears on the vita.¹⁵²

[87] Third, implicit biases are highly responsive to environmental exposure and experience.¹⁵³ We have already discussed studies relevant to *Grace's Hire*. Here, we shift gears to *Skip's Encounter*. Preference for White over Black (measured by the IAT) decreased following exposure to a positive depiction of Black Americans (a segment from the film *Poetic Justice*) than following exposure to a negative depiction (a clip of the film *Black & White & Red All Over* showing Black characters arguing over a gang-related incident).¹⁵⁴ Similarly, racial attitudes were less biased on a priming procedure when Black faces were viewed in a “church” context than an “urban street corner” context.¹⁵⁵

[88] Such findings raise the possibility of purposeful training with counter-typical datasets. For instance, repeated exposure to a dataset that associates Black faces with positive words have been decreased implicit bias, as measured by priming instruments.¹⁵⁶ The implicit stereotype associating Blacks with athleticism can similarly be dissipated, especially among participants highly motivated not to be biased.¹⁵⁷ And, even shooter bias is attenuated with practice among both police officer and non-police samples.¹⁵⁸

[89] In conclusion, the strict determinism component of “resignation” is rejected by the evidence, in cases that range from *Grace's Hire* to *Skip's*

¹⁵¹ See Kang & Banaji, *Fair Measures*, *supra* note 2, at 1092-1101 (outlining other strategies that individuals and firms could take to help disrupt the causal chain).

¹⁵² For other potential interventions, see Russell R. Robinson, *Perceptual Segregation*, 108 COLUM. L. REV. 1093, 1171-74 (discussing how diversifying workplace structures might help debias).

¹⁵³ See generally Irene V. Blair, *The Malleability of Automatic Stereotype*, 6 Pers. & Soc. Psych. Rev. 242 (2002) (literature review).

¹⁵⁴ See B. Wittenbrink et al., *Spontaneous prejudice in context: variability in automatically activated attitudes*, 81 J. PERSONALITY & SOC. PSYCHOL. 815, 818-19 (2001).

¹⁵⁵ *Id.* at 823.

¹⁵⁶ Similarly, repeated pairings of Black faces with positive words during an ostensibly unrelated exercise resulted in more egalitarian implicit racial attitudes, even though participants were unaware of any systematic pairing between positive words and Black faces. This reduction persisted for two days following exposure to the Black-positive pairing. See MA Olson & Russell H. Fazio, *Reducing automatically activated racial prejudice through implicit evaluative conditioning*, 32 PERSONALITY AND SOCIAL PSYCHOLOGY BULLETIN 421 (2006).

¹⁵⁷ See B. Michelle Peruche & E. Ashby Plant, *Racial bias in perceptions of athleticism: The role of motivation in the elimination of bias*, 24 SOCIAL COGNITION 438 (2006) (debiasing via repeated exposure to pairings of Black and White faces with athletic or nonathletic objects, where race and athletic features were independent).

¹⁵⁸ Racial shooting bias in a police simulation was decreased after repeated exposure to pairs of stimuli where ethnicity was unrelated to criminality. See EA Plant et al., *Eliminating automatic racial bias: Making race non-diagnostic for responses to criminal suspects*, 41 J. EXPERIMENTAL SOCIAL PSYCHOLOGY 141 (2005).

Encounter.¹⁵⁹ The law cannot ignore the behavioral realist call to account by throwing up its hands in resignation.

E. Justification

[90] One could argue, without apology, that implicit biases accurately reflect reality. For example, if we associate Asian faces more than White faces to the stereotype “foreign,” it is because that association reflects basic immigration realities. This helps explain why Asian American Yale students demonstrated this bias as strongly as their White peers.¹⁶⁰ If we associate women more with family than with work, again that is because the association reflects basic socioeconomic realities and division of labor within family units. This helps explain why women seem to hold this bias even stronger than men. Accordingly, “bias” is just a judgmental label for accurate probabilities, and preventing action based on them amounts to suppressing behavior based on the cold hard facts. This mode of response is the accuracy *justification*. If resignation was wistful, justification is defiant.

[91] As applied to explicit biases, this argument is familiar. Dinesh D’Souza for example defended the cab driver who declined to pick up the African American youth:

How hollow it sounds to accuse cabdrivers of “prejudices” and “stereotypes” when their perceptions seem to be based on empirical reality. While we can be sure that racist taxi drivers would discriminate, it is not clear that all taxi drivers who discriminate are racist...African American males have a right to be concerned about their convenience and dignity, but cabdrivers too are entitled to care about their property and safety.¹⁶¹

[92] What’s new here is the claim that not only are explicit biases accurate but implicit biases are as well. For instance, Gregory Mitchell and Philip Tetlock write:

The belief that stereotypes generally lead to erroneous judgments about targets may itself be an erroneous stereotype about stereotypes . . . For example, using the race of a person approaching on a dark street in a high crime area to predict that person’s criminal propensity may be

¹⁵⁹ Of course, one could soften the *strict* determinism claim to weaker forms of determinism. And if one were to insist that there is weak determinism (however defined), then there would be some justification for weak resignation (again however defined). But the flipside of weak resignation is strong determination to make a change.

¹⁶⁰ See Devos & Banaji, *supra* note 50.

¹⁶¹ DINESH D’SOUZA, THE END OF RACISM: PRINCIPLES FOR A MULTIRACIAL SOCIETY 252-53 (1995).

rational if one's race-based stereotypes reflect true differences in base rates of criminality across racial groups.¹⁶²

[93] The accuracy *justification* comprises two claims. The descriptive claim contends that the strength of associations in our brains accurately reflects the real state of the world.¹⁶³ The normative claim contends that acting on the basis of accurate pictures of the world is morally and legally justified.¹⁶⁴

1. Descriptive

[94] The descriptive claim of accuracy itself comprises two sub-claims: we carry accurate probabilities in our heads (the *accurate probability data* claim), and we act on the basis of those probabilities rationally (the *rational processing* claim).

a. Accurate probability data

[95] As Charles Judd and Bernadette Park have clarified, there are at least three different and potentially unrelated accuracy measures to consider: stereotypic accuracy (is there over or under-estimation of some perceived attribute as compared to the "real" measure?), valence accuracy (does this over or under-estimation depend on whether the trait is positive or negative?), and dispersion accuracy (is there an over or under-generalization that misestimates the variance of the attribute?).¹⁶⁵

[96] There are many reasons to be skeptical about claims of *stereotypic accuracy*. First, for this to be even plausible, the individual's implicit biases cannot be dissociated from his explicit ones; otherwise, one set of biases (sic probabilities) would have to be mistaken. Second, where do these base rates come from? In the domain of race, they come more through vicarious than direct experiences with racial others,¹⁶⁶ and there's little reason to think that we are fed accurate

¹⁶² Mitchell & Tetlock, *Mindreading*, *supra* note 120, at 1036 n.41. See also Hal Arkes and Philip Tetlock, who suggest that mental associations measured as implicit social cognitions may in fact be "accurate statistical associations rather than unwarranted conclusions." Arkes & Tetlock, *Jesse Jackson*, *supra* note 91, at 258

¹⁶³ Mitchell & Tetlock, *Mindreading*, *supra* note 120, at 1085 ("implicit prejudice, as now conceived, labels perfectly rational reactions to existing socioeconomic conditions as prejudiced").

¹⁶⁴ See *id.* at 1087 ("the courts permit employers to base decisions on job-relevant attributes correlated with protected-category membership"); *id.* at 1087-88 ("Just as it would be bizarre to constrain employers to base their decisions solely on variables that have zero correlations with membership in protected groups, it would be bizarre to expect people to fail to notice real-world statistical relationships involving protected categories-and to expect them not to form mental models of the world (associative networks) that reflect those relationships.").

¹⁶⁵ See Charles M. Judd & Bernadette Park, *Definition and Assessment of Accuracy in Social Stereotypes*, 100 PSYCH. REV. 109, 110-11 (1993).

¹⁶⁶ For a discussion of direct versus vicarious experiences with racial others, see Jerry Kang, *Cyber-race*, 113 HARVARD L. REV. 1131, 1166-67 (2000) ("By vicarious, I mean imagined experiences - both fictional and nonfictional - that are mediated through stories told by parents, teachers, friends, and

information through popular culture.¹⁶⁷ Third, we see illusory correlations, which arise when two salient events (minority + negative event) are noticed together, which leaves a larger impression on our memories and prompt us to overestimate the frequency of their connection.¹⁶⁸ Fourth, although the memory studies are conflicted, we seem to have preferential recording and recall of stereotype-consistent data.¹⁶⁹

[97] As for *valence accuracy*, we have substantial motivational desires to exaggerate good traits of ingroups and bad traits of outgroups.¹⁷⁰ Finally, as for *dispersion accuracy*, again, there will be predictable ingroup-outgroup errors. We tend to engage in outgroup homogenization, in which we think outgroups are more monolithic than the ingroups that we are more familiar with. Thus, we focus more on mean attributes and underestimate variance.¹⁷¹

[98] Finally, even though it makes sense to ask whether a stereotype is accurate, can we ask the same of an attitude? Attitudes are fundamentally evaluative, about liking/disliking, about having positive feelings/negative feelings. If one has a positive attitude toward cilantro in part because of early childhood exposure, is that accurate or inaccurate? If one has a negative attitude toward Mexicans—and who knows the reasons why?—is that somehow accurate? If one shows a preferential attitude towards one's ingroup—even if randomly assigned to made-up teams¹⁷²—again, in what ways is that accurate?

b. Rational processing

[99] Next, let's turn to the rational processing sub-claim. It is only a slight exaggeration to suggest that fifty years of hard-nosed scientific inquiry belies any claim of strict rationality.¹⁷³ The rich Behavioral Law & Economics literature has documented the science in the law reviews. We are, in fact, terrible calculators of probabilities and our behaviors do not seem rationally tailored to

increasingly by the electronic mass media. By contrast, direct experiences are actual experiences with people of other races, not mediated by a third party such as the mass media.” (footnote omitted).

¹⁶⁷ See Kang, *Trojan Horses*, *supra* note 20, at 1563-64; Robert M. Entman & Andrew Rojecki, *THE BLACK IMAGE IN THE WHITE MIND* 49 (2000) (suggesting that vicarious experiences dominate the construction of mainstream culture).

¹⁶⁸ See Kang, *Trojan Horses*, *supra* note 20, at 1564.

¹⁶⁹ See Judd & Park, *supra* note 165, at 112 (describing tendencies to overly attend to confirming evidence and underuse disconfirming evidence).

¹⁷⁰ See Judd & Park, *supra* note 165, at 112-13.

¹⁷¹ See Kang, *Trojan Horses*, *supra* note 20, at 1565.

¹⁷² See *supra* note 38.

¹⁷³ Amos Tversky and Daniel Kahneman, *Judgments under uncertainty: Heuristics and biases*, 185 *SCIENCE* 1124 (1974); Alen Newell & Herbert A. Simon, *HUMAN PROBLEM SOLVING* (1972). See Nosek, Banaji, & Greenwald, *Nostalgia*, *supra* note 91, at 285-86 (describing ambiguity of the term “rational” and explaining why individuals are likely to lack the necessary data and computation abilities to act rationally).

any such computations.¹⁷⁴ Moreover, our motivations to justify the self and the groups we belong to slant how we use or fail to use base-rate information.¹⁷⁵ Also, our interpretations of even accurate probability calculations will reflect biases. Consider, for example, the ultimate attribution error (which is related to the well-known fundamental attribution error), which makes us view negative attributes of outgroups as stable, fixed, and dispositional.¹⁷⁶ By contrast, negative attributes of ingroups are viewed as malleable, contingent, and due to environment or bad luck. Everything is flipped for positive attributes.

[100] Further, empirical evidence suggests that these biases can alter behavior even in ways that are self-detrimental. Eugene Caruso and colleagues¹⁷⁷ had participants choose partners in a trivia game based on traits that were relevant to success (such as IQ, prior experience) and traits irrelevant to success (such as weight, as signaled by a photograph). By using conjoint analysis,¹⁷⁸ the researchers were able to estimate that although participants reported that weight was “was the single least important factor in their choice”¹⁷⁹ a clear preference emerged for the thin partner over the overweight one. In fact, the influence of weight was not significantly different from IQ and prior experience. In quantifiable terms, participants sacrificed between 10.53 and 12.31 IQ¹⁸⁰ points to work with the thinner teammate.¹⁸¹ In another study with similar methodology, undergraduates were willing to trade away \$3,249 (22% of the salary range of options) to work for a man, instead of a woman.¹⁸² Leaving IQ points or money

¹⁷⁴ See, e.g., Aron K. Barbey & Steven A. Sloman, *Base-rate Respect: From Ecological Rationality to Dual Processes*, 30 BEHAVIORAL & BRAIN SCIENCES 241, 251-52 (2007) (providing summary of empirical findings on Bayesian reasoning).

¹⁷⁵ See Eric Luis Uhlmann, Victoria L. Brescoll, & David Pizarro, *The Motivated Use and Neglect of Base Rates*, 30 BEHAVIORAL AND BRAIN SCIENCES 284 (2007).

¹⁷⁶ See Thomas F. Pettigrew, *The Ultimate Attribution Error: Extending Allport's Cognitive Analysis of Prejudice*, 5 PERSONALITY & SOC. PSYCHOL. BULL. 461 (1979). See also Kang, *Trojan Horses*, *supra* note 20, at n.418 (describing phenomenon).

¹⁷⁷ See Eugene M. Caruso, Dobromir A. Rahnev, & Mahzarin Banaji, *Using Conjoint Analysis to Detect Discrimination: Revealing Covert Preferences from Overt Choices* 6-7 (2009) (describing method).

¹⁷⁸ For an introduction to conjoint analysis, see Bryan K. Orme, *GETTING STARTED WITH CONJOINT ANALYSIS: STRATEGIES FOR PRODUCT DESIGN AND PRIVACY RESEARCH* (2006), available at <<http://www.sawtoothsoftware.com/education/techlist2.shtml>>.

¹⁷⁹ See Caruso et al., *Conjoint Analysis*, *supra* note , at 133.

¹⁸⁰ The choices were made in two possible ways: evaluating potential partners one at a time (IQ points sacrificed: 10.53) or choosing between a pair of potential partners (IQ points sacrificed: 12.31) See *id.* at 134.

¹⁸¹ Now, this was a *much* thinner teammate, which pre-testing estimated at about 97 lbs heavier (at M=237 lbs) than the thinner teammate (at M=140 lbs). See *id.* at 134. If the effect is linear, which may or may not be the case, this translates to 1.27 IQ points for each 10 lbs of weight.

¹⁸² See *id.* at 136 (citing Rahnev et al., *Conjoint Analysis: A New method of Investigating Stereotypes*, 2007, unpublished manuscript).

on the table, especially when one explicitly reports that weight and gender don't matter, is hardly rational.¹⁸³

2. Normative

[101] Even if the descriptive accuracy burden can be discharged, there is a second round of questions that are normative. They can be normative in the following thin way. All measurements are subject to error, so when anyone claims that something is "accurate", she is really saying that something is "accurate enough." The word "enough" does enormous normative work, and embeds within it value judgments that correspond to what scientists call Type I (illusion) and Type II (blindspot) errors.¹⁸⁴ "Accurate enough" in *Skip's Encounter*—shoot or not shoot—is not the same question as "accurate enough" on a friend's pizza order—pepperoni or mushroom. Nothing about rationality tells us how we should evaluate the significance of these Type I and Type II errors.

[102] Even thicker normative anxieties are lurking. Procedurally, if we are making judgments based on probabilities of uncertain accuracy and moral significance, we should at the least publish these probabilities for both perceiver and target to appreciate.¹⁸⁵ For example, if a police officer believes that on the basis of accurate stereotypes, he should draw his gun more quickly on Skip (than Harvey, a comparably dressed White man), then shouldn't these data be publicly articulated and vetted? There is no such transparency today.

[103] Substantively, for many, even if the stereotype is descriptively "accurate," it is normatively wrong to treat individuals differently on the basis of that stereotype in many contexts. This intuition is shared not only among some radical Left: it's repeated by scholars such as Steven Pinker in his popular rejection of the *Blank Slate*.¹⁸⁶

[104] Finally, we should remind ourselves where the law is on this "accuracy" point. Negative treatment based on what seems like a probabilistically accurate stereotype is often illegal. Indeed, as Samuel Bagenstos writes: "The prohibition of rational discrimination is a central component of antidiscrimination doctrine—and it may be the most important aspect of antidiscrimination law on

¹⁸³ Again, an apologist could suggest that one's true preferences have been revealed by the conjoint analysis, so the behavior *was* rational. At this point, the faith in rationality has become nearly unfalsifiable.

¹⁸⁴ A Type I error is a false positive, which involves rejecting the null hypothesis when it is in fact true. This is seeing something that is not actually there. A Type II error is a false negative, which involves accepting the null hypothesis when it is in fact false. This is being blind to something that is actually there.

¹⁸⁵ See Nosek, Banaji, & Greenwald, *Nostalgia*, *supra* note 91, at 285.

¹⁸⁶ See STEVEN PINKER, *THE BLANK SLATE: THE MODERN DENIAL OF HUMAN NATURE* 145 (2002) ("It is a moral stance that condemns judging an individual according to the average traits of certain groups to which the individual belongs.").

the ground.”¹⁸⁷ Accordingly, to contend that the law need not account for implicit biases and their actuated behavior because of accuracy, one must not only make the empirical case of accuracy but also the normative case why so much of current ethical and legal understandings ought to be overturned. That isn’t easy.¹⁸⁸

CONCLUSION

[105] Once upon a time, the central civil rights questions were indisputably normative. What did “equal justice under law” require? Did it, for example, permit segregation, or was separate never equal? This is no longer the case. Today, the central civil rights questions of our time turn also on the underlying empirics. In a post-civil rights era, in what some people exuberantly embrace as a post-racial era, many assume that we already live in a color- and gender-blind society. The only exceptions are either hate criminals or universities that insist on affirmative action. In between these extremes, the rest of “us” have learned well from Martin Luther King, Jr. and now judge people only on the content of their characters, not by their social categories. Accordingly, if some groups underperform in various economic, political, and educational competitions, those disparities evince not injustice but incompetence. And in modern capitalist societies, “we” are not responsible for the fate of the incompetent “them” except for mild tax-and-transfer subsidies.

[106] This convenient story is, however, disputed. In the past, within the legal literature, it was disputed by passionate retelling of lived experiences.¹⁸⁹ But these narratives were resisted as mere anecdotes, not much different than the purely fictional vignettes of Juan, Skip, and Grace. In the past, the convenient story was disputed with sociological analyses of structures, institutions, and cultures. But this was the “societal discrimination” that the Supreme Court rejected in *Wygant* saying that “[s]ocietal discrimination, *without more*, is too amorphous a basis for imposing a racially classified remedy” because a “court could uphold remedies that are ageless in their reach into the past, and timeless in their ability to affect the future.”¹⁹⁰

¹⁸⁷ See Samuel R. Bagenstos, “Rational Discrimination,” *Accommodation, and the Politics of (Disability) Civil Rights*, 89 VA. L. REV. 825, 848 (2003). See also Samuel R. Bagenstos, *The Supreme Court, the Americans with Disabilities Act, and Rational Discrimination*, 55 ALA. L. REV. 923, 935-45 (2004).

¹⁸⁸ This is a descriptive claim of what the current law is, and what the accuracy justification must surmount. It is not a normative claim that because the law is what it is, it should stay that way. Now, as a normative matter, we agree that action based on probabilistically accurate enough stereotypes should in many cases still be proscribed. But we confess that we haven’t made that normative argument here in any detail.

¹⁸⁹ See, e.g., PATRICIA J. WILLIAMS, *THE ALCHEMY OF RACE AND RIGHTS: DIARY OF A LAW PROFESSOR* (1991).

¹⁹⁰ *Wygant v. Jackson Bd. of Educ.*, 476 U.S. 267, 276 (1986) (Powell, J.) (plurality opinion) (emphasis added).

[107] Now, the convenient story is also being contested with *something more*—the modern authority of empirical evidence from the mind sciences. We now have accumulated hard data, collected from scientific experiments, with all their mathematical precisions, objective measurements, and statistical dissections. And this scientific knowledge has the potential to burst our complacent assumptions of objectivity and fairness. It can fundamentally alter how we might accept responsibility for the inequities around us.

[108] In this Article, we have presented a sketch of the science of implicit social cognition and how a commitment to behavioral realism could guide the science's incorporation into the law writ large. As with any new branch of scientific knowledge, there are uncertainties and controversies, which should be rigorously examined. To repeat, on-the-merits challenges sounding in *denial* and *minimization* are to be encouraged, not suppressed. But it is naïve to think that only the merits will determine the science's uptake. Accordingly, we must be mindful of *politicization* across the political spectrum. And when the scientific consensus arrives, we should expect the modes of response sounding in (passive) *resignation* and (aggressive) *justification*.

[109] Rather than recapping our analyses of each of these modes, in this Conclusion, we demonstrate a different sort of payoff of this grandly gestured future history. We believe that this zoomed-out and longitudinal projection allows us to see otherwise obscured interrelations. We highlight only three.

[110] First, notice that *minimization* conflicts with *resignation*. Minimization contends that implicit biases have at most *de minimis* impact. By contrast, resignation claims that implicit biases strictly determine our behavior. It's hard for both to be true.

[111] Second, notice that the accuracy *justification* conflicts with both *denial* and *minimization*. *Denial* suggests that there is no such thing as implicit bias and that the measuring instruments, such as the IAT, lack validity. *Minimization* suggests that even if implicit biases exist, they do not manifest in real world behaviors. In sharp contrast, the accuracy justification supposes that implicit biases exist and are created by the fact that we have senses that respond to external reality and brains that recognize patterns. Moreover, those who push this objection often believe that humans are relatively accurate expected utility calculators. If I have accurate probabilities in my head, then of course I will act on them in the real world.¹⁹¹ But how can this view be reconciled with *minimization*?

[112] Finally, notice the inconsistent burdens of proof invoked. During *politicization*, the backlash from the Right called the science junk, and demanded validities to the nines: statistical validity, internal validity, construct validity,

¹⁹¹ Cf. Arkes & Tetlock, Jesse Jackson, *supra* note , at 270 ("The general principle is this: As long as there is a differential crime rate between racial groups, a perfectly rational decision maker may manifest different behaviors—explicit and implicit—toward members of different races.").

predictive validity, external validity, ecological validity, and still other flavors. Shouldn't the same evidentiary standard be consistently required of the accuracy *justification*?

[113] In the end, only hindsight is 20/20, and no doubt this future history will have made mistakes. That said, by providing clear framing and names, we will have clarified the terms of engagement and surfaced hidden interconnections. At the least, we will have accelerated the substantive debate past caricatures. It is said that those who do not remember the past are condemned to repeat it. We add that those who can imagine the future can build it.

Table 1

Task	Positive values indicate	N	IAT			Self-report		
			Mean	Standard deviation	d	Mean	Standard deviation	d
Age attitude	preference for young people compared to old people	351,204	0.49	0.39	1.23	0.39	0.78	0.51
Race attitude	preference for Whites compared to Blacks	732,881	0.37	0.43	0.86	0.26	0.73	0.36
Skin-tone attitude	preference for Light-skin compared to Dark-skin	122,988	0.30	0.41	0.73	0.17	0.67	0.25
Child-race attitude	preference for White children compared to Black Children	28,816	0.33	0.45	0.73	0.19	1.30	0.15
Arab-Muslim attitude	preference for other people compared to Arab-Muslims	77,254	0.14	0.42	0.33	0.45	0.77	0.58
Judaism attitude	preference for other religions compared to Judaism	66,092	7.15	0.44	7.34	0.14	1.05	0.13
Disability attitude	preference for abled people compared to disabled people	38,544	0.45	0.43	1.05	0.38	0.67	0.57
Sexuality attitude	preference for straight people compared to gay people	269,683	0.35	0.47	0.74	0.49	0.91	0.54
Weight attitude	preference for thin people compared to fat people	199,329	0.35	0.42	0.83	0.64	0.73	0.88
Race-Weapons stereotype	stronger Black Weapons, White=Harmless objects associations than the reverse	85,742	0.37	0.37	1.00	0.34	1.10	0.31
American-Native stereotype	stronger White American=American, Native American=Foreign associations than the reverse	44,878	0.23	0.50	0.46	7.76	1.79	7.42
American-Asian stereotype	stronger European American=American, Asian American=Foreign associations than the reverse	57,569	0.26	0.41	0.62	0.57	1.27	0.45
Gender-Science stereotype	stronger Male=Science, Female=Humanities associations than the reverse	299,298	0.37	0.40	0.93	0.52	0.66	0.79
Gender-Career stereotype	stronger Male=Career, Female=Family associations than the reverse	83,084	0.39	0.36	1.10	0.54	0.60	0.89
Presidential attitude	preference for Bush compared to other US presidents	68,123	7.07	0.45	7.15	7.94	1.28	7.73
Election 2004 attitude	preference for Bush compared to Kerry	22,904	7.14	0.51	7.27	7.69	1.64	7.42
Election 2000 attitude	preference for Bush compared to Gore	27,146	7.09	0.56	7.16	7.32	1.60	7.20

Adapted from Table 1 (pp. 38-39) and Table 2 (p. 46) in Brian A. Nosek et al., *Pervasiveness and correlates of implicit attitudes and stereotypes*, 18 European Review of Social Psychology, 36 (2007).

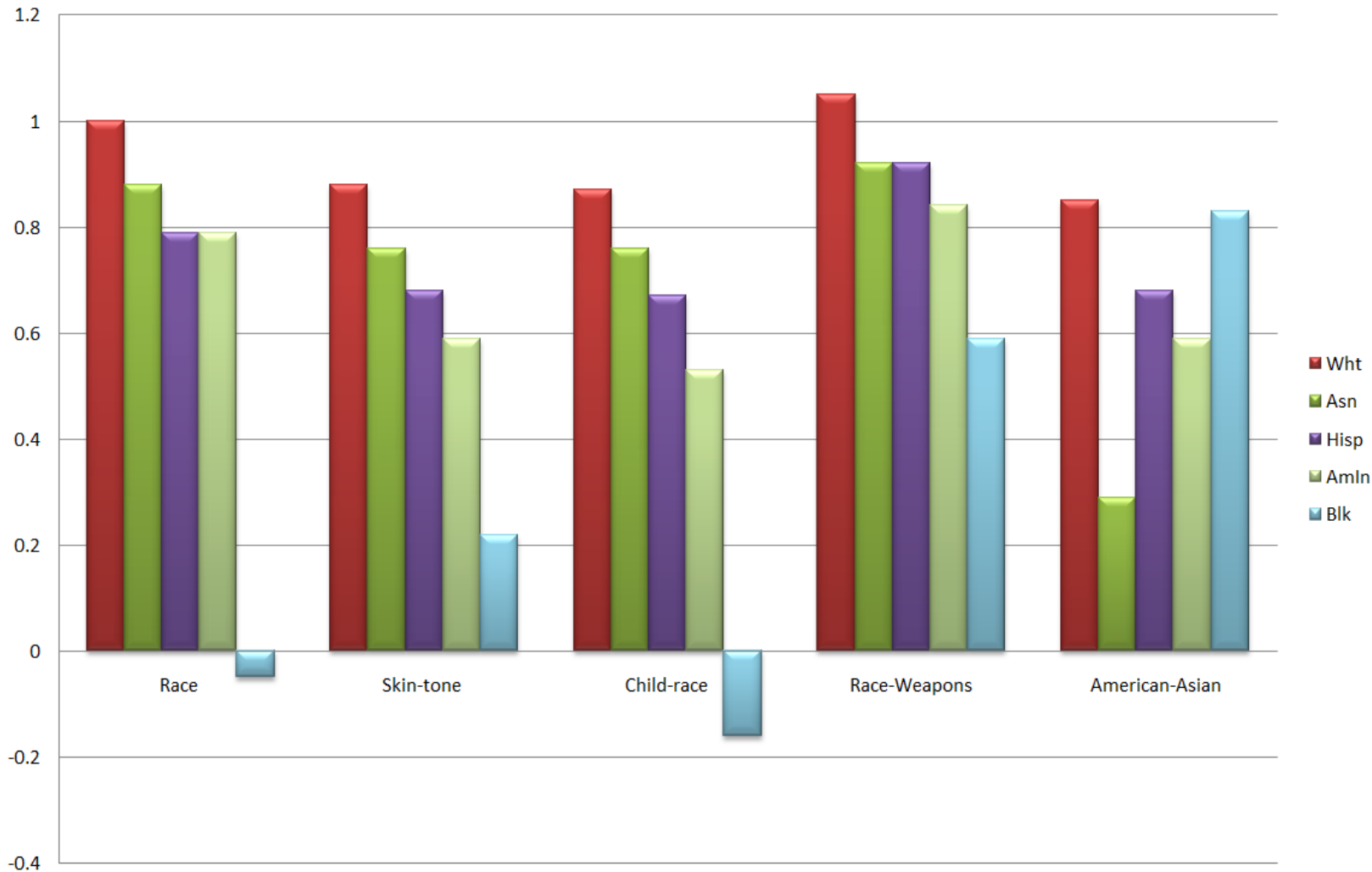
N = number of completed IATs. IAT means are D scores as calculated per the description in Anthony G. Greenwald et al., *Understanding and using the Implicit Association Test: I. An improved scoring algorithm*, 81, J. Personality and Social Psychol. 197 (2001).

Explicit means represent the mean for a subset of self-report measures used.

d represents Cohen's d, a standardized unit of the size of a statistical effect. By convention, .20, .50 and .80 represent small, medium and large effect sizes, respectively. (Jacob Cohen (1988). Statistical Power Analysis for the Behavioral Sciences (second ed.). Lawrence Erlbaum Associates.)

The reported means for self-reported attitudes or stereotypes are for a selected item or comparison of items.

Figure 1: Implicit Biases by Race



Adapted from data presented in Table 4 (pp. 48) in Brian A. Nosek et al., *Pervasiveness and correlates of implicit attitudes and stereotypes*, 18 *European Review of Social Psychology*, 1 (2007). The Figure depicts IAT scores for four different tasks by self-reported ethnicity. Scores on the y-axis represent Cohen's *d* scores (described in Table 1). Whites, Asians, Hispanics, and Native Americans showed, on average, robust preference for White over Black (adults and children) and Light Skin over Dark Skin. Black participants' implicit preferences were closer to the neutral point, with slight preference for Black over White (adults and children), and Light Skin over Dark Skin. People of all races tended to associate *Black* and *weapons* (compared to *White* and *Harmless objects*), although this tendency was somewhat muted among Black test-takers. Similarly, people of all races tended to associate *Asian* with *Foreign*, although this tendency was muted with Asians.