SEEING THROUGH COLORBLINDNESS: IMPLICIT BIAS AND THE LAW

Jerry Kang*

Kristin Lane**

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 post-racial, many assume that we already live in a colorblind society. Is this in fact the case? Recent findings about implicit bias from mind scientists sharply suggest otherwise. This Article summarizes the empirical evidence that rejects facile claims of perceptual, cognitive, and behavioral colorblindness. It then calls on the law to take a "behavioral realist" account of these findings, and maps systematically how it might do so in sensible, nonhysterical, and evidence-based ways. Recognizing that this call may be politically naive, the Article examines and answers three objections, sounding in "junk science" backlash, "hardwired" resignation, and "rational" justification.

---

* Professor of Law, UCLA School of Law. kang@law.ucla.edu. http://jerrykang.net.

** Assistant Professor of Psychology, Bard College. Financial disclosure statement: The authors have earned honoraria from academic lectures on implicit bias. Professor Kang also serves as a paid consultant to the National Center for State Courts' project on implicit bias.

For helpful comments, we thank Gary Blasi, Devon Carbado, Rachel Godsil, Carol Goldberg, Anthony Greenwald, Cheryl Harris, Ken Karst, Sung Hui Kim, Doug Kysar, Greg Mitchell, Jennifer Mnookin, Jeff Rachlinski, Russell Robinson, Len Rubinowitz, Ed Stein, Michael Sullivan, Steven Willborn, and Noah Zatz. Special thanks go to Mahzarin Banaji. We also acknowledge our research assistants: Jonathan Feingold, Jin Goh, Rita Halpert, Kevin Minnick, Vivek Mittal, Jennie Park, Emily Reitz, Yi (Jenny) Xiao, and Hentyle Yapp. We also benefitted from the prompt research performed by the Hugh & Hazel Darling Law Library at UCLA School of Law.

Earlier incarnations of this Article were presented at: Seton Hall Faculty Colloquium; Harvard Public Law Seminar; Emory Legal Theory Workshop; Yale Work Seminar; Southwestern University School of Law; Conference of Asian Pacific American Law Professors 2008 Keynote, Denver Law School; Berkeley Law School's Thelton Henderson Center Conference "Reclaiming & Reframing the Dialogue on Race and Racism"; Ohio State University School of Law; UCLA Critical Race Studies Faculty Seminars; Northwestern Critical Race Theory Colloquium.

This project was supported in part by resources from the UCLA School of Law, UCLA Asian American Studies Center, UCLA Institute for American Cultures, and the Korea Times-*Hankook Ilbo* Chair in Korean American Studies.

INTRODUCTION

We start with three fictional stories.

Juan's Interview. Juan was desperate for the job. But he didn't get it. Crushed, Juan grabs a beer with his friend. Juan grouses 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." 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." Juan is a Chicano.

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, he jimmies it open. Just as he lays down his bags, exhausted, he notices 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 sick and tired. 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 turns pale, as he stares down the barrel of a gun. He stammers slowly, "it's my faculty ID . . . ."[1]

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 with a high Asian American population, 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 that Grace is good but that Ben is 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 probably illegal. Ben happens to be White and liberal, like most of the faculty. He gets hired; Grace does not.

How do you respond to these stories? No doubt your reactions turn on your values. For example, is colorblindness a moral or legal imperative—even in *Grace's Hire*? But your reactions also depend crucially on the facts. What really happened in Juan's interview? Who's really to blame in Skip's testy encounter? Would Grace's hire really "do us good," or is that just wishful thinking? Put another way, does race really matter? Even when we put aside difficult philosophical questions about equality, justice, and fairness, we still run into tough empirical questions of whether we are, in fact, "colorblind."

Thankfully, there has been a recent explosion of scientific knowledge on this 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] These implicit thoughts and feelings leak into everyday behaviors such as whom we befriend, whose work we value, and whom we favor—notwithstanding our

1. To repeat, this is fiction. Most important, when Professor Skip Gates was arrested at his home, the arresting officer did not draw his gun. For a journalistic account, see Abby Goodnough, *Harvard Professor Jailed; Officer Is Accused of Bias*, N.Y. TIMES, July 21, 2009, at A13.

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

obliviousness to any such influence. These discoveries, disseminated from hundreds of research programs, have provoked surprise, fascination, and even anger.

As behavioral realists, we believe such findings 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.[3] Theories and data from the mind sciences 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 it should provide reasons why it cannot or will not do so.

Part I introduces the science of implicit social cognition and makes the scientific case against colorblindness: We are not perceptually, cognitively, or behaviorally colorblind.[4] With this background, Part II articulates a call for behavioral realism and maps out how social and legal institutions might respond to the new discoveries. Our goal here is not to argue in painstaking detail in favor of specific legal or policy reforms; instead, our goal is to schematize conceptually the various ways that the law could incorporate the new science. Part III answers three predictable objections: "junk science" backlash, "hardwired" resignation, and "rational" justification. Throughout the Article, we revisit *Juan's Interview*, *Skip's Encounter*, and *Grace's Hire*, to see how the mind sciences simultaneously complicate and clarify our simplistic presumptions about colorblindness.

I. SEEING THROUGH COLORBLINDNESS

A. The Psychological Context of Colorblindness

Juan's interviewer believes that he is colorblind. Skip's arresting officer believes that he is colorblind. The faculty members who voted for Ben and not Grace believe that they are colorblind. But how well founded are such assumptions?

As a threshold matter, we must unpack the term "colorblind." To be perceptually colorblind is not to even see race in the way comedian Stephen Colbert insists he does not.[5] There is, however, simply too much evidence of

3. *See generally infra* Part II.A.

4. Although the same underlying cognitive and social processes apply to other social categories such as gender, nationality, and age, our focus is on race. In other words, the more general phenomenon of interest is category blindness or category agnosticism. But we stick with the more familiar example of colorblindness.

5. The character Stephen Colbert of *The Colbert Report* regularly reminds his viewers of his perceptual colorblindness. He knows he's White only because other people tell him so. This seems to be

automatic classification of individuals into social categories, including race, to maintain this position. To the contrary, race and ethnicity are highly salient and chronically accessible categories.[6] Thus, when people claim colorblindness, they cannot be claiming perceptual colorblindness; instead, they are likely claiming to be cognitively colorblind. In other words, they have no (meaningfully) different attitudes or stereotypes between any two racial categories (e.g., Black and White).[7]

We purposefully distinguish attitude and stereotype. For social psychologists, an attitude is "an association between a given object and a given evaluative category."[8] 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 Blacks" is an example of an attitude, consciously reported.[9] By contrast, a stereotype does not emphasize an evaluative association as much as a more specific attribute, such as "Latinos are undocumented."[10]

These social cognitions, whether they be attitudes or stereotypes, can be either implicit or explicit. Similar to its usage in cognitive psychology, the term "implicit" emphasizes our unawareness of having a particular thought or feeling.[11] 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 articulate having such

---

one of the few appropriate occasions to cite Wikipedia. *See Stephen Colbert (character)*, WIKIPEDIA, http://en.wikipedia.org/wiki/Stephen_Colbert_(character) (last visited Nov. 11, 2010).

6. Race (and other social group memberships such as age and sex) appears to be encoded with no substantial effort on the perceiver's part. *See* Shelley E. Taylor et al., *Categorical and Contextual Bases of Person Memory and Stereotyping*, 36 J. PERSONALITY & SOC. PSYCHOL. 778, 782–83 (1978) (showing that race is an organizing principle used in social interactions). Indeed, the very act of trying to be colorblind and ignore race seems to deplete cognitive resources in a way that can actually impair interracial interactions. Evan P. Apfelbaum et al., *Seeing Race and Seeming Racist? Evaluating Strategic Colorblindness in Social Interaction*, 95 J. PERSONALITY & SOC. PSYCHOL. 918, 924–25 (2008) (showing that people who tried to avoid thinking about race performed worse on a task of cognitive control following an interracial interaction).

7. A third way to understand colorblindness is in terms of behavioral colorblindness. We discuss that version *infra* Part I.B.2.

8. *E.g.*, Russell H. Fazio et al., *Attitude Accessibility, Attitude-Behavior Consistency, and the Strength of the Object-Evaluation Association*, 18 J. EXPERIMENTAL SOC. PSYCHOL. 339, 341 (1982).

9. A negative attitude is often called prejudice.

10. Of course, a particular stereotype may support an overall evaluative attitude. If one dislikes undocumented people and one associates Latinos with undocumented status, then this attribute will likely contribute to a negative attitude toward Latinos.

11. 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."). Regarding 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 8 n.2.

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

We have attitudes and stereotypes, both explicit and implicit, about objects, like beef, 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, which isn't controversial. But when applied to humans, these basic concepts have been sharply protested.[12] We understand that it is more unsettling to hold two separate (and often divergent) attitudes toward a social group, such as Asians, as compared to mere objects, such as beef or broccoli. But our anxiety says nothing about whether differential implicit social cognitions in fact exist regarding Whites versus Asians.[13] What legal and policy analysts must appreciate is that the findings in nonsocial and social domains rest on identical scientific foundations.

So, are folks in fact cognitively colorblind? To answer this question, we could simply ask people like Juan's interviewer for their honest self-reports about their attitudes and stereotypes. But this methodology would be naive for two reasons.[14] First, respondents engage in impression management and provide what they feel are politically correct answers.[15] Second, introspection is strikingly undependable. As Richard Nisbett and Timothy Wilson reported over three decades ago, "[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

12. For instance, Hal Arkes and Philip Tetlock suggest 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*." Hal R. Arkes & Philip E. Tetlock, *Attributions of Implicit Prejudice, or "Would Jesse Jackson 'Fail' the Implicit Association Test?"*, 15 PSYCHOL. INQUIRY 257, 268 (2004) (emphasis added). But numerous mental constructs, such as attention, perception, and memory all have explicit and implicit counterparts. We do not reject the concept of implicit memory simply because the memory cannot be consciously recalled. Accordingly, we see no reason to carve out attitudes and stereotypes for disparate treatment.

13. We do have implicit biases against Asians, and these biases do predict behavior, such as our evaluations of their lawyering. *See* Jerry Kang et al., *Are Ideal Litigators White? Measuring the Myth of Colorblindness*, 7 J. EMPIRICAL LEGAL STUD. (forthcoming Dec. 2010).

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

15. 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. PERSONALITY & SOC. PSYCHOL. 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").

16. Richard E. Nisbett & Timothy DeCamp Wilson, *Telling More Than We Can Know: Verbal Reports on Mental Processes*, 84 PSYCHOL. REV. 231, 233 (1977). An updated version of the argument appears in Timothy D. Wilson & Elizabeth W. Dunn, *Self-Knowledge: Its Limits, Value, and Potential for Improvement*, 55 ANN. REV. PSYCHOL. 493, 494 (2004) (discussing the "several reasons why people are not an open book to themselves").

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]

To overcome these difficulties, cognitive and social psychologists have developed instruments to measure cognitions without asking for self-reports.[18] Some tests use linguistic cues;[19] others use physiological instruments, for instance, by measuring cardiovascular responses,[20] micro-facial movements,[21] or neurological activity.[22] But the largest class of measures relies on reaction times when completing various tasks. The basic assumption of these reaction-time 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," we are quicker to identify "nurse" as a word than we are to identify "purse" as a word. By measuring the speeds of activations, we can infer their strength of association—in this example, concluding that there is a relatively stronger association between "doctor" and "nurse" than between "doctor" and "purse." As simplistic as this approach sounds, reaction-time instruments[23] have produced the most reliable measures for implicit social cognitions.[24]

---

17. Nisbett & Wilson, *supra* note 16, at 257.

18. *See* Russell H. Fazio & Michael A. Olson, *Implicit Measures in Social Cognition Research: Their Meaning and Use*, 54 ANN. REV. PSYCHOL. 297–327 (2003) (reviewing various implicit measures in general and semantic priming and the Implicit Association Test in particular).

19. *See, e.g.*, Denise Sekaquaptewa et al., *Stereotypic Explanatory Bias: Implicit Stereotyping as a Predictor of Discrimination*, 39 J. EXPERIMENTAL SOC. PSYCHOL. 75 (2003); Lisa Sinclair & Ziva Kunda, *Reactions to a Black Professional: Motivated Inhibition and Activation of Conflicting Stereotypes*, 77 J. PERSONALITY & SOC. PSYCHOL. 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) (investigating whether mere description versus explanation is proffered to describe a behavior).

20. *See* Jim Blascovich et al., *Perceiver Threat in Social Interactions With Stigmatized Others*, 80 J. PERSONALITY & SOC. PSYCHOL. 253, 253–57 (2001).

21. *See* Eric J. Vanman et al., *Racial Discrimination by Low-Prejudiced Whites: Facial Movements as Implicit Measures of Attitudes Related to Behavior*, 15 PSYCHOL. SCI. 711, 711 (2004).

22. *See* Jason P. Mitchell et al., *Thinking About Others: The Neural Substrates of Social Cognition*, *in* SOCIAL NEUROSCIENCE: PEOPLE THINKING ABOUT THINKING PEOPLE 63, 64–65 (John T. Cacioppo et al. eds., 2006); Elizabeth A. Phelps et al., *Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation*, 12 J. COGNITIVE NEUROSCI. 729, 729 (2000).

23. 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 in Recognizing Pairs of Words: Evidence of a Dependence Between Retrieval Operations*, 90 J. EXPERIMENTAL PSYCHOL. 228 (1971). This design was later modified to study evaluative priming—the extent to which a concept is associated with "good" and "bad." *See* Russell H. Fazio et al., *On the Automatic Activation of Attitudes*, 50 J. PERSONALITY & SOC. PSYCHOL. 229, 229 (1986) [hereinafter Fazio, *Automatic Activation*]. In this measure, pairs of items

Among these instruments, the most widely used is the Implicit Association Test (IAT).[25] The IAT requires participants to rapidly classify individual stimuli into one of four distinct categories using only two responses (for example, participants might respond using only the "E" key on the left side of a computer keyboard, or "I" on the right side).[26] For instance, in a race attitude IAT, there are two social categories, EUROPEAN AMERICAN and ASIAN AMERICAN, and two attitudinal categories, GOOD and BAD. EUROPEAN AMERICAN and ASIAN AMERICAN 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 a positive or negative effect, such as *Joy* or *Agony*.[27] A person with a negative implicit attitude toward Asian Americans would be expected to go more quickly when ASIAN and BAD share one key, and EUROPEAN AMERICAN and GOOD the other, than when the pairings of GOOD

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, 1015–16 (1995) (describing the first adaptation of the evaluative priming procedure to measure automatic racial attitudes) [hereinafter Fazio, *Variability*]. Each word is 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.* at 1013.

24. *See supra* note 18.

25. Seminal work in this domain was performed by Russell Fazio, who explored how priming could help measure automatic attitudes and beliefs. *See* Bernd Wittenbrink, *Measuring Attitudes Through Priming*, *in* IMPLICIT MEASURES OF ATTITUDES 17 (Bernd Wittenbrink & Norbert Schwarz eds., 2007). Newer tasks, such as the Go/No-Go Association Task, *see* Brian A. Nosek & Mahzarin R. Banaji, *The Go/No-Go Association Task*, 19 SOC. COGNITION 625, 626–31 (2001), the Evaluative Movement Assessment, *see* C. Miguel Brendl et al., *Indirectly Measuring Evaluations of Several Attitude Objects in Relation to a Neutral Reference Point*, 41 J. EXPERIMENTAL SOC. PSYCHOL. 346, 347 (2005), the extrinsic affective Simon task, *see* Jan De Houwer, *The Extrinsic Affective Simon Task*, 50 EXPERIMENTAL PSYCHOL. 77, 79–80 (2003), and the affect misattribution procedure, *see* B. Keith Payne, *An Inkblot for Attitudes: Affect Misattribution as Implicit Measurement*, 89 J. PERSONALITY & SOC. PSYCHOL. 277, 277–79 (2005), have been used, and likely more will be developed. *See* Fazio, *Automatic Activation*, *supra* note 23; Fazio, *Variability*, *supra* note 23, at 1015–16.

26. *See* Anthony G. Greenwald et al., *Measuring Individual Differences in Implicit Cognition: The Implicit Association Test*, 74 J. PERSONALITY & SOC. PSYCHOL. 1464, 1464–66 (1998) (introducing the IAT).

27. In the racial attitude IAT, available at Project Implicit, 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 et al., *Harvesting Implicit Group Attitudes and Beliefs From a Demonstration Web Site*, 6 GROUP DYNAMICS: THEORY, RESEARCH, AND PRACTICE 101, 114 (2002).

and BAD are switched.[28] The time difference, which is recalibrated into what is known as the IAT *D* score, is interpreted as reflecting an implicit attitude.[29]

B. The Implicit Social Cognitive Challenge

1. Against Cognitive Colorblindness

For brevity's sake, we focus on the Implicit Association Test (IAT), although other reaction-time instruments have produced consistent findings. Since 1998 when the IAT was officially introduced, hundreds[30] of peer-reviewed scientific publications have produced largely consistent results. Implicit biases—by which we mean implicit attitudes and stereotypes—are both pervasive (most individuals show evidence of some biases), and large in magnitude, statistically speaking. In other words, we are not, on average or generally, cognitively colorblind.

Clear evidence of the pervasiveness of implicit bias comes from Project Implicit,[31] a research website operated by Harvard University, Washington University, and the University of Virginia. At Project Implicit, visitors can try IATs that examine implicit attitudes and stereotypes on topics ranging across race, gender, age, politics, region, religion, and even consumer brands. With over seven million completed tests, Project Implicit comprises the largest available repository of implicit social cognition data.[32]

---

28. 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, *supra* note 25, at 59; Brian A. Nosek, Anthony G. Greenwald, & Mahzarin R. Banaji, *The Implicit Association Test at Age 7: A Methodological and Conceptual Review*, *in* AUTOMATIC PROCESSES IN SOCIAL THINKING AND BEHAVIOR 265 (John A. Bargh ed., 2007), *available at* http://www.projectimplicit.net/nosek/iat/.

29. *See* Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm*, 85 J. PERSONALITY & SOC. PSYCHOL. 197, 200–01 (2003) (providing an improved IAT scoring method). There is an implicit attitudinal preference in favor of White males over East Asian males. *See* Kang et al., *supra* note 13, at ¶ 57 (finding a negative implicit attitude against Asian Americans) (IAT D=0.62, $t(67)$=13.31, p<0.001).

30. A PsycINFO database search for "implicit association test" appearing anywhere in the results, conducted in September 2010, yielded 2613 published works, with 2039 of them in peer-reviewed journals. *See* screen captures (on file with author) (performed on Sept. 9, 2010, 3:27 PM).

31. PROJECT IMPLICIT, http://projectimplicit.org (last visited Oct. 7, 2010).

32. 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 a web-based sample (from volunteers across the globe) broadly converge with laboratory data. Additionally, the enormous size of the data repository from the internet allows questions to be answered with confidence that is not otherwise possible. For a general discussion about the benefits and costs of using internet data, see Nosek et al., *supra* note 27, at 102–04.

Table 1 provides results from the seventeen IATs available at Project Implicit.[33] Most participants demonstrated implicit attitudes in favor of one social group over another, away from the neutral position of no bias. Notwithstanding protestations to the contrary, people are generally not "color" 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.[34]

TABLE 1[35]

			IAT			Self-Report		
Task	**Positive values indicate**	**N**	**Mean**	**Std. dev'n**	***d***	**Mean**	**Std. dev'n**	***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 WHITE PEOPLE compared to BLACK PEOPLE	732,881	0.37	0.43	0.86	0.26	0.73	0.36
Skin-tone attitude	Preference for LIGHT-SKIN PEOPLE compared to DARK-SKIN PEOPLE	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	–0.15	0.44	–0.34	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

33. *See* Brian A. Nosek et al., *Pervasiveness and Correlates of Implicit Attitudes and Stereotypes*, 18 EUR. REV. SOC. PSYCHOL. 1, 3–4 (2007).

34. *See id.* at 11.

35. *Id.* at 36 (adapted from tbl.1 (pp. 38–39) and tbl.2 (p. 46)). N=number of completed IATs. IAT means are *D* scores as calculated per the description in Greenwald et al., *supra* note 29. Explicit means represent the mean for a subset of questions on which participants reported their preferences or stereotypes; *d* represents Cohen's *d*, a standardized unit of the size of a statistical effect. By convention, 0.20, 0.50, and 0.80 represent small, medium, and large effect sizes, respectively. JACOB COHEN, STATISTICAL POWER ANALYSIS FOR THE BEHAVIORAL SCIENCES (2d ed. 1988). The means for self-reported attitudes or stereotypes are for a selected item or comparison of items.

TABLE 1 (Continued)

			IAT			Self-Report		
Task	**Positive values indicate**	**N**	**Mean**	**Std. dev'n**	***d***	**Mean**	**Std. dev'n**	***d***
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	–0.76	1.79	–0.42
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	–0.07	0.45	–0.15	–0.94	1.28	–0.73
Election 2004 attitude	Preference for BUSH compared to KERRY	22,904	–0.14	0.51	–0.27	–0.69	1.64	–0.42
Election 2000 attitude	Preference for BUSH compared to GORE	27,146	–0.09	0.56	–0.16	–0.32	1.60	–0.20

These hierarchy-driven biases sometimes counter expected ingroup favoritism[36]—our tendency to favor the groups we belong to.[37] Accordingly, those who belong to social groups deemed to be "good" (e.g., the young, European Americans, straight people) show strong preference for their own group. On the other hand, those who come from groups that the culture assigns as "bad" (e.g., the elderly, African Americans, gay people) do not show strong ingroup preference. Often their data show no ingroup preference at all and sometimes even tilt in the opposite direction.[38] In laboratory studies, we see similar findings for participants who belong to lower-status social groups marked by weight, age, sexual orientation, socioeconomic class,[39] university status (i.e. students at lower-ranked universities),[40] and college dorm status (i.e. students at less popular dorms).[41] In other words, on implicit measures, the overweight, poor, and those associated with less-valued institutions all showed weaker preference for their own group.

If we shift focus away from implicit attitudes to implicit stereotypes, the question changes from liking or disliking to whether some social category is

36. *See* Henri Tajfel & John C. Turner, *The Social Identity Theory of Intergroup Behavior*, *in* POLITICAL PSYCHOLOGY: KEY READINGS IN SOCIAL PSYCHOLOGY 276, 280–81 (John T. Jost & Jim Sidanius eds., 2004).

37. For example, on IATs, Japanese Americans and Korean Americans preferred their own ethnic group relative to the other, *see* Greenwald, *supra* note 26, at 1471–72, as did East and West Germans, *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, 139 (2001). Ingroup bias is so strong that people explicitly report liking "ingroups" even when they are randomly assigned to them. *See* Maria Rosaria Cadinu & Myron Rothbart, *Self-Anchoring and Differentiation Processes in the Minimal Group Setting*, 70 J. PERSONALITY & SOC. PSYCHOL. 661, 661–62 (1996); Samuel L. Gaertner et al., *Reducing Intergroup Bias: The Benefits of Recategorization*, 57 J. PERSONALITY & SOC. PSYCHOL. 239, 239 (1989); Henri Tajfel et al., *Social Categorization and Intergroup Behaviour*, 1 EUR. J. SOC. PSYCHOL. 149, 151 (1971). This is the case even when the groups are made up. For example, participants were told they preferred an unknown (and fictitious) artist 'Quan' or 'Xanthie' and that people who preferred this artist tended to have a particular kind of information-processing style. With this small and evaluatively meaningless amount of information, participants nevertheless showed implicit biases in favor of their assigned group. *See* Leslie Ashburn-Nardo et al., *Implicit Associations as the Seeds of Intergroup Bias: How Easily Do They Take Root?*, 81 J. PERSONALITY & SOC. PSYCHOL. 789, 795–98 (2001).

38. *See* Leslie Ashburn-Nardo et al., *Black Americans' Implicit Racial Associations and Their Implications for Intergroup Judgment*, 21 SOC. COGNITION 61, 73 (2003); Robert W. Livingston, *The Role of Perceived Negativity in the Moderation of African Americans' Implicit and Explicit Racial Attitudes*, 38 J. EXPERIMENTAL SOC. PSYCHOL. 405, 411–12 (2002). Africans in the United States and in South Africa show substantially weaker ingroup preference. *See* Kristina 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, N.M. (2008).

39. *See* Laurie A. Rudman et al., *Minority Members' Implicit Attitudes: Automatic Ingroup Bias as a Function of Group Status*, 20 SOC. COGNITION 294, 311–13 (2002).

40. *See* John T. Jost et al., *Non-Conscious Forms of System Justification: Implicit and Behavioral Preferences for Higher Status Groups*, 38 J. EXPERIMENTAL SOC. PSYCHOL. 586, 592 (2002).

41. *See* Kristin A. Lane et al., *Me and My Group: Cultural Status Can Disrupt Cognitive Consistency*, 23 SOC. COGNITION 353, 380–81 (2005).

linked to some attribute. For instance, the question isn't whether one likes or dislikes men versus women (an attitude);[42] 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 participants (72 percent) associated the concepts MALE with SCIENCE and FEMALE with HUMANITIES.[43] Similarly, participants found it easier to categorize White faces, rather than Asian American or Native American faces, with AMERICAN, reflecting an implicit stereotype that "American = White."[44] Regardless of what might be the case on the explicit level, on the implicit level, we are not cognitively colorblind.

But what if these measures are a horrible technical mistake caused by instrumentation or mathematical error? Maybe we are colorblind after all, and the machines simply got us wrong. Good scientists instinctively respond to any new measurement device with skepticism, and reaction-time measures such as the IAT have been no exception.[45] In particular, scientists have carefully examined both its reliability[46] and its validity.[47] After a decade of research, we believe that the IAT has demonstrated enough reliability and validity that total denial is implausible.

As for reliability, across twenty studies, the average (and median) correlation between a person's IAT score at two different times was 0.50,[48] which is a respectable psychometric measure.[49] Because of the moderate reliability, nearly

42. In an important counterexample of ingroup favoritism, "men are less likely than women to show automatic ingroup bias (i.e., own gender preference)." 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.

43. Nosek et al., *supra* note 33, at 21.

44. *See* Thierry Devos & Mahzarin R. Banaji, *American = White?*, 88 J. PERSONALITY & SOC. PSYCHOL. 447, 452–53, 457 (2005).

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

46. 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 it might be insufficiently reliable for someone trying to lose five pounds.

47. 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 are no confounds?"), and construct validity ("Are you sure you're actually measuring what you think you're measuring?").

48. *See* Lane et al., *supra* note 28, at 70 (aggregating test-retest reliabilities across twenty such administrations).

49. This result means that IAT scores at two time points share approximately 25 percent of their variance (0.50-squared). A person's score fluctuates around some mean. Pearson's *r*, a correlation coefficient, is always a number ranging from –1.0 to 1.0, which quantifies the strength of the linear

all scientists have discouraged using the IAT in high-stakes individual selection contexts, such as judicial nominations.[50] But this level of reliability is perfectly adequate for the IAT's efficacy as a research tool because we can aggregate across people to discern general patterns between mental constructs and behavior.[51]

As for validity, questions about statistical conclusion validity have been addressed.[52] So have potential confounds that would undermine internal validity, including selection,[53] familiarity of the stimuli,[54] the order in which the

relationship between two variables (in this case, IAT scores at two time points). Each correlation coefficient provides information about the direction (by the correlation's sign) and strength (by the correlation's magnitude) of the relationship between the two variables. For explanations of the statistical concepts, see ALAN AGRESTI & BARBARA FINLAY, STATISTICAL METHODS FOR THE SOCIAL SCIENCES (2d ed. 1986). By squaring the *r*, we get the percentage of variance explained, which reveals the extent to which differences among people in a population can be attributed to a single variable. In this example, we see that r^2=0.25. This value means that we can account for 25 percent of the variability in one of a person's IAT scores by using the score from the other time point as a linear predictor. To offer another example, if we knew that the correlation between a father's height and a child's height was 0.70 (this number is illustrative for this purpose), when looking at a group of people—some short, some average, some tall—49 percent (=0.7 x 0.7) of the differences in height among them could be accounted for by knowing their fathers' heights. (Notably, more than half of the variation in heights among them would be due to other factors, such as diet, mother's height, other genetic influences, and luck.).

50. *See* Shankar Vedantam, *See No Bias*, WASH. POST, Jan. 23, 2005, at W12 (reporting that Mahzarin Banaji and her colleagues "will testify in court against any attempt to use the test to identify biased individuals").

51. One final counterintuitive point: Lower reliability levels (which reflect added noise) result in *underestimation* of causal relationships; so, to the extent that our analysis errs, it does so by understating the impact of implicit bias. *See* C. Spearman, *The Proof and Measurement of Association Between Two Things*, 15 AM. J. PSYCHOL. 72 (1904).

52. 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* Hart Blanton & James Jaccard, *Arbitrary Metrics in Psychology*, 61 AM. PSYCHOLOGIST 27, 33–34 (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 et al., *Consequential Validity of the Implicit Association Test: Comment on Blanton and Jaccard*, 61 AM. PSYCHOLOGIST 56, 58–59 (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).

53. *See*, *e.g.*, Jason P. Mitchell et al., *Contextual Variations in Implicit Evaluation*, 132 J. EXPERIMENTAL PSYCHOL.: GEN. 455, 467–68 (2003) (finding that when items changed the categories' construal, implicit biases were diminished). However, analyses of large web-based datasets suggest that unless items 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 Association Test II: Method Variables and Construct Validity*, 31 PERSONALITY & SOC. PSYCHOL. BULL. 166, 170–71 (2005).

54. *See* Nilanjana Dasgupta et al., *The First Ontological Challenge to the IAT: Attitude or Mere Familiarity?*, 14 PSYCHOL. INQUIRY 238, 239–42 (2003). Dasgupta et al. argued that three lines of evidence

combined tasks are completed,[55] previous experience with the IAT,[56] inter-trial interval duration,[57] assignment of categories to a right or left key,[58] right- or left-handedness of the participant,[59] cognitive fluency of the participant,[60] and fakeability.[61] Finally, there are answers to construct validity[62] challenges

---

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). *Id.* at 241. 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. *Id.* at 238–43; *see also* Nilanjana Dasgupta et al., *Automatic Preference for White Americans: Eliminating the Familiarity Explanation*, 36 J. EXPERIMENTAL SOC. PSYCHOL. 316, 325–26 (2000); Scott A. Ottaway et al., *Implicit Attitudes and Racism: Effects of Word Familiarity and Frequency on the Implicit Association Test*, 19 SOC. COGNITION 97, 130 (2001).

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 in the laboratory. *See* Ashburn-Nardo et al., *supra* note 37, at 789.

55. *See* Greenwald et al., *supra* note 26; Nosek et al., *supra* note 53, at 177–79 (finding that order has a small effect on IAT scores, which can be minimized by increasing the number of practice trials).

56. *See* Greenwald et al., *supra* note 29, at 211 (finding that prior experience slightly reduced the magnitude of IAT effects in subsequent trials).

57. *See* Greenwald et al., *supra* note 26, at 1469 (finding essentially no effect on the magnitude of IAT depending on the time interval, ranging from one-hundred milliseconds (ms) (or one-tenth of a second) to 700 ms, between successive trials).

58. *See id.* (finding no effects). The IAT counterbalances the side to which, for instance, pleasant and unpleasant categories are assigned by systematically varying the lateral location of the categories between participants.

59. *See* Anthony G. Greenwald & Brian A. Nosek, *Health of the Implicit Association Test at Age 3*, 48 ZEITSCHRIFT FÜR EXPERIMENTELLE PSYCHOLOGIE 85, 87 (2001) (finding no effects of self-reported handedness in a large web-based dataset).

60. *See* Sam G. McFarland & Zachary Crouch, *A Cognitive Skill Confound on the Implicit Association Test*, 20 SOC. COGNITION 483, 503–06 (2002); Jan Mierke & Karl Christoph Klauer, *Method-Specific Variance in the Implicit Association Test*, 85 J. PERSONALITY & SOC. PSYCHOL. 1180, 1189–90 (2003). Using improved scoring techniques that control for differences among participants in general response speed 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 SOC. COGNITION 673, 680–81 (2004).

61. *See, e.g.*, Klaus Fiedler & Matthias Bluemke, *Faking the IAT: Aided and Unaided Response Control on the Implicit Association Tests*, 27 BASIC & APPLIED SOC. PSYCHOL. 307, 314–15 (2005); Melanie C. Steffens, *Is the Implicit Association Test Immune to Faking?*, 51 EXPERIMENTAL PSYCHOL. 165, 175–76 (2004); Do-Yeong Kim, *Voluntary Controllability of the Implicit Association Test (IAT)*, 66 SOC. PSYCHOL. Q. 83, 95–96 (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 would undermine the basic results. Implicit measures often reveal attitudes and stereotypes that are quite discrepant from self-reported ones. Given that people want to minimize the appearance of bias—even implicit biases—faking test scores simply indicates that implicit biases in the population are in fact larger than reported.

suggesting that the IAT measures not individual bias but salience asymmetry[63] or general cultural associations unattributable to any specific person.[64]

This brief summary cannot exhaustively detail the evidence in support of the IAT's reliability and validity. Curious readers should turn to the cited

62. 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 them to converge with (e.g., other measures of attitudes and stereotypes) and depart from measures that we would expect them 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 responses, particularly fear—when viewing unfamiliar Black (versus White) faces. *See* William A. Cunningham et al., *Neural Components of Social Evaluation*, 85 J. PERSONALITY & SOC. PSYCHOL. 639, 640 (2003); Phelps et al., *supra* note 22. 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 that they had been presented) than when they were presented supraliminally (faces appeared on the screen long enough for participants to be aware of seeing them). 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 PSYCHOL. SCI. 991, 996–97 (2007).

63. *See* Klaus Rothermund & Dirk Wentura, *Underlying Processes in the Implicit Association Test: Dissociating Salience From Associations*, 133 J. EXPERIMENTAL PSYCHOL.: GEN. 139, 156 (2004). The claim here is that particularly salient items are easier to group together. 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 Implicit Association Test: Comment on Rothermund and Wentura (2004)*, 134 J. EXPERIMENTAL PSYCHOL.: GEN. 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 PSYCHOL.: GEN. 426 (2005).

64. *See*, *e.g.*, Arkes & Tetlock, *supra* note 12; Phillip E. Tetlock & Hal R. Arkes, *The Implicit Prejudice Exchange: Islands of Consensus in a Sea of Controversy: Response*, 15 PSYCHOL. INQUIRY 311–21 (2004); Michael A. Olson & Russell H. Fazio, *Reducing the Influence of Extrapersonal Associations on the Implicit Association Test: Personalizing the IAT*, 86 J. PERSONALITY & SOC. PSYCHOL. 653, 663–65 (2004); Andrew Karpinski & James L. Hilton, *Attitudes and the Implicit Association Test*, 81 J. PERSONALITY & SOC. PSYCHOL. 774, 786–87 (2001). For responses and analysis, see Mahzarin R. Banaji et al., *No Place for Nostalgia in Science: A Response to Arkes and Tetlock*, 15 PSYCHOL. INQUIRY 279, 283–85 (2004); Brian A. Nosek & Jeffrey J. Hansen, *The Associations in Our Heads Belong to Us: Searching for Attitudes and Knowledge in Implicit Cognition*, 22 COGNITION & EMOTION 553, 582–88 (2008); Eric Luis Uhlmann et al., *Automatic Associations: Personal Attitudes or Cultural Knowledge?*, *in* IDEOLOGY, PSYCHOLOGY, AND LAW (Jon Hanson ed., 2010), *available at* http://www.projectimplicit.net/articles.php. We resist overly stylized demarcations that separate culture from person. Quite obviously, culture plays a huge role in feeding the associations in our brains. *See*, *e.g.*, Kang, *supra* note 14, at 1539–40 (describing the role of mass media providing "vicarious" experiences with racial others). But if IAT scores simply reflected a tally of the associations seen in the world around us and 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, *supra*.

scientific papers, which include critiques and responses. More important, we are not claiming that the IAT is somehow immaculate. No psychological measure is perfectly pure; even a simple self-reported scale to assess a person's self-esteem[65] can be affected by variables such as participants' literacy, the comfort of the room in which they respond, and any immediately prior success or failure. We simply suggest that the IAT is no different than other psychological measures, and the evidence suggests that none of the potential confounds can account for IAT effects in the hundreds of studies that have used the measure.

2. Against Behavioral Colorblindness

All that we have reported so far are pervasive differences in a computerized speed test. But do these differences, measured in milliseconds, matter in the real world? In other words, even if a person is not cognitively colorblind, these cognitions may not affect real-world behavior,[66] in which case, who cares? Perhaps, cognitive colorblindness is unimportant; it's behavioral colorblindness that we're really after. We analyze this possibility with the help of our fictional stories.

Recall *Skip's Encounter*. Perhaps Skip could have defused the situation simply by putting on a smile. But it turns out that 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 than on the White face. Moreover, the extent that hostility was perceived as lingering was predicted by implicit bias (as measured by the IAT) against Blacks.[67]

If a neutral facial expression is seen as more angry on a Black face, could a wallet be mistaken for a gun more often when held by a Black man? Unfortunately, the data suggest yes. A "shooter bias" paradigm has explored whether people are prone to accidentally shoot Black suspects more often than White

65. For example, "On the whole, I am satisfied with myself" is an item from the Rosenberg self-esteem scale. *See* MORRIS ROSENBERG, SOCIETY AND THE ADOLESCENT SELF-IMAGE 307 (1965).

66. 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.

67. *See* Kurt Hugenberg & Galen V. Bodenhausen, *Facing Prejudice: Implicit Prejudice and the Perception of Facial Threat*, 14 PSYCHOL. SCI. 640, 641–42 (2003) (showing that when the video went from hostile to friendly, implicit anti-Black bias predicted the extent to which hostility would "linger"; p=0.04). Implicit bias scores predicted responses to Black β=0.46, p=0.02, but not White, β=0.09, *ns*, faces. *See id.* at 642. In a second study, the video showed faces going from friendly to hostile. In this situation, people with higher implicit bias scores were quicker to report detecting hostility in Black rather than White faces, p=0.02. *See id.*

suspects. In this measure, 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.[68]

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.[69] 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 armed Black targets.[70] Interestingly, these stereotype-consistent behaviors emerged among both Black and White participants.[71] On the margins, then, *Skip's Encounter* was more dangerous because of his race, regardless of the cop's race. These findings suggest that we don't find behavioral colorblindness even in deadly serious situations.

But maybe a police confrontation is too extreme an example (especially for readers who experience police interactions as courteous and nonthreatening). What about the more mundane context of *Juan's Interview*? Why did that interview go 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.[72] In one of the earliest demonstrations of

68. Joshua Correll et al., *The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals*, 83 J. PERSONALITY & SOC. PSYCHOL. 1314, 1315–17 (2002) (describing the test procedure); *see also* Anthony G. Greenwald et al., *Targets of Discrimination: Effects of Race on Responses to Weapons Holders*, 39 J. EXPERIMENTAL SOC. PSYCHOL. 399, 400–01 (2003) (finding similar results).

69. *See* Correll et al., *supra* note 68, at 1317.

70. *See id.* at 1319.

71. *See id.* at 1324–25.

72. *See* 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, 438, 441 (2001).

Skeptics have challenged McConnell and Leibold's results. *See* Hart Blanton et al., *Strong Claims and Weak Evidence: Reassessing the Predictive Validity of the IAT*, 94 J. APPLIED PSYCHOL. 567, 574–76 (2009). Specifically, Blanton et al. argue that the relationship between the IAT and the judges' global ratings of the interactions rely too heavily on unreliable ratings by the judges and are dependent on a single judge's ratings. We concur with McConnell and Leibold, who in their reply echo the point we made earlier: "[A]ny dissimilarities between judges' ratings would only increase variability in their assessments, making it more (not less) difficult to observe the significant relations between the IAT and biased behaviors found in the study." Allen R. McConnell & Jill M. Leibold, *Weak Criticisms and Selective Evidence: Reply to Blanton et al.*, 94 J. APPLIED PSYCH, 583, 585 (2009).

The skeptics also raise concerns about a participant who was several decades older than other participants and exhibited an implicit preference for WHITE several standard deviations greater than the sample mean, without whose data the correlation between the IAT and observers' ratings of participants no longer met conventional standards of significance (r=0.34, p<0.05), but would still be considered "marginally" significant (r=0.27, p <0.10). Blanton et al., *supra*, at 572–73.

implicit racial attitudes, people with more negative implicit attitudes toward Blacks (compared to Whites) on an evaluative priming measure were rated less friendly by a Black interaction partner who was unaware of their scores than those with more positive implicit attitudes toward Blacks.[73] This unfriendliness seeps out through nonverbal cues: the more negative the implicit attitude, the more awkward the body language.

Body language matters because if one person acts awkwardly to another, the other person will reciprocate, thus generating a vicious circle.[74] This interaction will then sour the interview without either party recognizing the implicit causal forces. If Juan's interviewer had a negative implicit attitude toward Latinos, that

---

We share McConnell and Leibold's puzzlement that they would exclude a participant with an admittedly large IAT score based on age but would retain another participant with an even higher IAT score in their reanalysis. *See* McConnell & Leibold, *supra*, at 583–84.

Perhaps most importantly, while the critique focuses on the relationship between implicit bias and a single behavior, McConnell and Leibold rightly point out that considering the full universe of measured behaviors demonstrates the robustness of their original findings:

> [A]s people revealed relatively more negativity toward Blacks on the IAT, they had relatively more negative explicit attitudes toward Blacks; the White experimenters perceived relatively more positive interactions with the participants than did the Black experimenters; and the judges viewed that the participants showed less speaking time, less smiling, more speech errors, and fewer extemporaneous social comments toward the Black experimenter than the White experimenter. Each and every one of these findings is directly at odds with Blanton's conclusion that the IAT did not predict behavioral bias in interracial interactions once the outlier was dismissed. Thus, out of the seven significant correlations between the IAT and biased behavior (as assessed both by the experimenters and by the judges) reported in McConnell and Leibold, five remained significant at conventional levels following the elimination of the outlier. Moreover, three of the correlations were larger following the elimination of the outlier too. Thus, we see considerable evidence that the IAT continues to predict indicators of biased intergroup behavior even with the elimination of this outlier.

Id. at 584–85.

Blanton et al. responded that even this list does not encompass all of the criterion variables in the original McConnell and Leibold report. Hart Blanton et al., *Transparency Should Trump Trust: Rejoinder to McConnell and Leibold (2009) and Ziegert and Hanges (2009)*, 94 J. APPLIED PSYCHOL. 598, 599 (2009). By now, the scientific back-and-forth will have become tedious. While individual studies can and should be critiqued, every experiment will be somewhat imperfect. Qualitative and quantitative reviews of the entire literature offer a perspective of the totality of the evidence.

73. *See* Fazio et al., *Variability*, *supra* note 23, at 1019.

74. *See* 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); Carl O. Word et al., *The Nonverbal Mediation of Self-Fulfilling Prophecies in Interracial Interaction*, 10 J. EXPERIMENTAL SOC. PSYCHOL. 109, 119 (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).

attitude may have leaked through his body language, which could have triggered a negative response from Juan.[75]

If we zoom out on Juan's timeline, things get potentially worse for him. Before Juan got the interview, he had to submit a resume. In 2004, behavioral economists Marianne Bertrand and Sendhil Mullainathan sent comparable resumes out to numerous employers in Boston and Chicago but used names such as "Emily" or "Greg" to signal Whiteness and "Lakisha" or "Jamal" to signal Blackness.[76] They found that the trivial manipulation of name produced a 50 percent difference in callback rates.[77]

In 2007, Dan-Olof Rooth replicated and extended this experiment by connecting the callback discrimination to implicit bias. He created resumes that again were similar in background, experience, and qualifications but varied the applicant's name to denote ethnicity.[78] Applications from an Arab-Muslim job candidate (whose experience, education, and short biography made clear that he was born and educated in Sweden) and a Swedish job candidate were each sent to 1552 posted job listings in Sweden.

Only one of the two applicants was invited for an interview in 283 of the job postings. And in these cases, a clear preference emerged for the candidate with the Swedish-sounding name (217 invites) over the one with

75. 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 actively trying to mask bias. J. Nicole Shelton et al., *Ironic Effects of Racial Bias During Interracial Interactions*, 16 PSYCHOL. SCI. 397, 400–01 (2005).

Despite this complexity, we do not believe that implicit biases influence behavior randomly. One interesting study tracked individuals with panic disorder during a twelve-week treatment program. Panic and anxiety symptomology decreased over the course of treatment. *See* Bethany A. Teachman et al., *Automatic Associations and Panic Disorder: Trajectories of Change Over the Course of Treatment*, 76 J. COUNSELING & CLINICAL PSYCHOL. 988, 991, 993 (2007) (as measured by the Anxiety Sensitivity Scale, the Beck Depression Inventory II, the Fear Questionnaire-Agoraphobia subscale and the Panic Disorder Severity Scale). The researchers used dynamic latent growth models to explore the trajectory of changes in symptoms. Decreases in the association between ME and *Panicked* (compared to NOT ME and *Calm*), measured by the IAT, were correlated with improvements in symptoms, r=0.28. *Id.* at 994. Importantly, "change in automatic panic associations significantly predict[ed] change in panic symptoms, but the reciprocal relationship" was not statistically significant. *Id.* at 995. Changes in implicit cognitions occurred prior to—and predicted—improvement in psychological health. These findings offer credence to the hypothesis that changes in implicit cognitions translate into behavioral changes in the same direction.

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

77. *See id.* at 998.

78. *See* Dan-Olof Rooth, *Implicit Discrimination in Hiring: Real World Evidence* 5 (Inst. for the Study of Labor, Discussion Paper No. 2764, 2007).

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. Coupled with the cases in which neither or both applicants were invited, these data reveal that candidates with Swedish-sounding names had an approximately 9 percent advantage in callback rates.[79]

Several months later, Rooth located 193 of the recruiters or managers responsible for the hiring decision. They reported explicit attitudes and stereotypes in their hiring preferences[80] and completed an IAT assessing the Arab-Muslim male stereotypes. Recruiters generally reported some explicit preference for both hiring Swedish men (54 percent of recruiters) and having warmer feelings toward them (45 percent 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*.[81]

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 percent decrease in the probability that an Arab-Muslim candidate received an interview.[82] Put another way, approximately half of the large discrepancy 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.[83]

Let's tie this evidence back to *Juan's Interview*. His name alone on the resume might make it harder to get an interview because of implicit stereotypes. If Juan gets a callback, then the social interaction may sour, partly because of negative implicit attitudes. And if he were to get the job, managers may interpret ambiguous performance in ways that confirm prior expectations.[84] We do

79. *See id.* In 239 cases, both applicants were invited to interview.

80. They indicated strong, moderate, or some preference for hiring Arab-Muslim men (in Sweden) to native Swedish men. *Id.* at 8.

81. *See id.* at 8–9.

82. Among all firms, including those that had extended no interview offers, a one standard deviation increase was associated with a smaller (3 percent) 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 measures did not significantly predict decisions, although Rooth noted that such bias may nevertheless be "economically important." *Id.* at 13.

83. *See id.* at 16.

84. *See, e.g.*, John M. Darley & Paget H. Gross, *A Hypothesis-Confirming Bias in Labeling Effects*, 44 J. PERSONALITY & SOC. PSYCHOL. 20, 27–29 (1983). In this study, cues about a child's social class affected interpretation of her academic performance. Perceivers who believed she was from a higher

not want to overstate the case. If Juan is truly stellar, he will likely be celebrated; if Juan is truly incompetent, subtle biases are the least of his worries. But if he is somewhere in the vast undifferentiated middle, then implicit biases may substantially alter his work trajectory over the life of his career.[85]

One final objection might emerge from the organizational accountability literature, which suggests that in the real word, firms have adopted decision-making procedures to minimize the behavioral manifestation of various biases, including implicit ones.[86] We agree that implicit biases are malleable and that the environment can strongly influence how and whether implicit biases translate into behavior.[87] That said, it is heroic to think that minimizing implicit biases is easy[88] and is somehow already taking place.

Motivation to be egalitarian should influence whether implicit bias translates into discriminatory behavior.[89] Indeed, data confirm this. In one study, implicit bias as measured by priming techniques predicted differential trait ratings of Blacks (relative to Whites) for those who report little motivation to control bias; however, this was not the case with those highly motivated to control bias.[90] Similar results have been found with IAT-measured implicit bias.[91] In other

socioeconomic status described her performance and abilities as stronger than perceivers who believed she was from a lower socioeconomic status. *Id.*

85. It's also possible that increased social interaction on the job could help decrease implicit bias. This is what the social contact hypothesis suggests. *See, e.g.*, Kang & Banaji, *supra* note 2, at 1101–05 (summarizing social contact hypothesis findings); Thomas F. Pettigrew & Linda R. Tropp, *A Meta-Analytic Test of Intergroup Contact Theory*, 90 J. PERSONALITY & SOC. PSYCHOL. 751, 751–52 (2006).

86. *See, e.g.*, Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 OHIO ST. L.J. 1023, 1108–10 (2006) (listing twelve factors that explain why laboratory findings might not generalize to the real world).

87. For a general discussion of the debiasing literature, see, for example, Adam Benforado & Jon Hanson, *The Great Attributional Divide: How Divergent Views of Human Behavior Are Shaping Legal Policy*, 57 EMORY L.J. 311 (2008) (finding that intergroup contact negatively correlates with prejudice).

88. *See* John A. Bargh, *The Cognitive Monster: The Case Against the Controllability of Automatic Stereotype Effects*, *in* DUAL-PROCESS THEORIES IN SOCIAL PSYCHOLOGY 361, 376–78 (Shelly Chaiken & Yaacov Trope eds., 1999).

89. *See* Russell H. Fazio & Tamara Towles-Schwen, *The MODE Model of Attitude-Behavior Processes*, *in* DUAL-PROCESS THEORIES IN SOCIAL PSYCHOLOGY, *supra* note 88, at 97.

90. 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 & APPLIED SOC. PSYCHOL. 1, 4 (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 & SOC. PSYCHOL. BULL. 170, 179 (2003). Similarly, for White participants low in motivation to control prejudicial responses, implicit bias predicted anticipated comfort levels during an unscripted interaction with a Black partner. However, implicit bias was not predictive for those strongly motivated to control prejudice. *See id.* at 176–78.

91. For example, implicit prejudice toward gay people predicted nonverbal nondiscriminatory behavior during an interaction with a gay partner only for participants with egalitarian motives and a tendency to control their behavior. *See* Nilanjana Dasgupta & Luis M. Rivera, *From Automatic*

words, one can be cognitively color conscious and still be behaviorally colorblind. We won't be motivated to check bias, however, if we are cocksure that we are bias-free to begin with. Unfortunately, we readily see bias in others but not in ourselves.[92]

This analysis cross-applies to institutions. To be sure, the accountability literature reveals that individuals who must explain their decisionmaking to others are less prone to various biases.[93] And, in the workplace, today's human resources departments and supervisors are accountable to their superiors for employment actions. That said, organizations generally do not ask their decisionmakers to account for implicit-bias-actuated discrimination. Given our misguided faith in our own objectivity[94] and the difficulty identifying whether implicit bias was a but-for (much less proximate) cause for any isolated behavior, it seems premature to assume that general accountability pressures are automagically solving[95] the problem.[96]

Indeed, the evidence is to the contrary. Recall *Juan's Interview*. If accountability to antidiscrimination laws and policies were solving the problem, "Jamal" should have been called back as often as "Greg." Recall *Skip's Encounter*. Surely our brave men and women in blue feel heightened accountability to avoid shooting an innocent person. Although the evidence is mixed, 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.[97]

Antigay Prejudice to Behavior: The Moderating Role of Conscious Beliefs About Gender and Behavioral Control, 91 J. PERSONALITY & SOC. PSYCHOL. 268, 270 (2006).

92. Emily Pronin et al., *The Bias Blind Spot: Perceptions of Bias in Self Versus Others*, 28 PERSONALITY & SOC. PSYCHOL. BULL. 369, 369–70 (2002). For a review, see Emily Pronin, *How We See Ourselves and How We See Others*, 320 SCIENCE 1177 (2008).

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

94. *See generally* David Alain Armor, The Illusion of Objectivity: A Bias in the Perception of Freedom From Bias (1998) (unpublished PhD dissertation, UCLA) (on file with author); 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) (on file with author).

95. *See* Philip E. Tetlock & Gregory Mitchell, *Implicit Bias and Accountability Systems: What Must Organizations Do to Prevent Discrimination?*, 29 RES. ORG. BEHAV. 3, 16–31 (2009).

96. *See, e.g.*, Lerner & Tetlock, *supra* note 93, at 270 (warning readers against believing "that accountability is a cognitive or social panacea . . . [that] '[a]ll we need to do is hold the rascals accountable'"). We thus agree more with Tetlock collaborating with Lerner in 1999 than with Tetlock collaborating with Mitchell in 2006. *See* Mitchell & Tetlock, *supra* note 86, at 1120–21.

97. E. Ashby Plant & B. Michelle Peruche, *The Consequences of Race for Police Officers' Responses to Criminal Suspects*, 16 PSYCHOL. SCI. 180, 181–82 (2005). In another study, police officers showed a similar tendency to respond faster to armed Blacks (compared to armed Whites) and unarmed Whites (compared to unarmed Blacks), but they 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. PERSONALITY & SOC. PSYCHOL. 1006, 1010–13, 1015–17 (2007) (describing results from two studies).

Rather than slog through more predictive validity studies, we rely on two summaries. In a recent paper, John Jost and colleagues[98] respond to critiques of the implicit bias science by Philip Tetlock and Gregory Mitchell.[99] Jost et al. note that summary dismissal of the research on implicit bias would be tantamount to disputing the "major tenets of 20th century cognitive psychology."[100] They review "10 studies that no manager should ignore" showing the effects of implicit bias in various domains, including hiring, policing, budget cuts, verbal slurs, medical diagnoses, and voting.[101]

If ten studies aren't enough, how about 122? Anthony Greenwald and colleagues completed a meta-analysis of 122 research reports that included 184 independent samples and 14,900 subjects, and that included the IAT and participant behaviors in domains ranging from personal relationships to consumer preferences to intergroup biases.[102]

This meta-analysis revealed that both explicit biases (measured by self-reports on surveys) and implicit biases (measured by the IAT) predicted certain variables, such as nonverbal behavior, social judgments, physiological responses, and social action. The correlations for each bias were moderate in size, with explicit biases being a slightly better predictor on average across all topics, including issues such as voting behavior and consumer choices.[103] However, the IAT predicted socially sensitive behavior *better* than did explicit self-reports.[104]

In the White-Black discrimination domain (i.e. whether anti-Black implicit attitude predicts behavior toward Blacks), implicit bias predicts 5.7

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

99. *See* Tetlock & Mitchell, *supra* note 95, at 16–31.

100. Jost et al., *supra* note 98, at 46.

101. *Id.* at 49–52.

102. Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity*, 97 J. PERSONALITY & SOC. PSYCHOL. 17, 19–20 (2009). A meta-analysis stitches together the weighted findings of all 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 is not confounded by such faults unless they appear systematically throughout the dataset, which is far less likely.

103. Across all topic areas and behaviors, the average IAT-behavior correlation was r=0.274, whereas the average correlation between explicit measures and behavior was r=0.361. *Id.* at 28.

104. The implicit bias correlations for White-Black discrimination were average r=0.24, significantly higher than explicit bias correlations of average r=0.12. Implicit bias also predicted behavior better in other intergroup discriminations dealing with other ethnic groups, age, and weight. In these cases, implicit biases correlated with behavior on average r=0.20, whereas explicit biases correlated on average r=0.12. *See id.* at 24 tbl.3. In the domain of gender and sexual orientation, explicit measures (r=0.22) were better predictors than the IAT (r=0.18).

percent of the variability in behavior.[105] One might retort that this number is too low to have policy significance.[106] But the correlation between implicit bias and behavior may be substantially higher, and may be masked by imperfections in the instruments that measure both the bias and the behavior.[107] Even if this is not the case, explaining 5.7 percent of the variability in Black-White discrimination is meaningful because we are summing up all the differential treatment over all interactions, over all time, and over all the people within each racial category.

Here's a baseball analogy. With two outs in the ninth inning, and a man on second base, who do you want at the plate: All-Star Matt Holliday (0.321 average in 2008) or John Buck (0.224 average in 2008)? In statistical terms, Robert Abelson has demonstrated that the batting average difference explains only 1.3 percent of the variance in a single at-bat (hit or no hit).[108] 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.[109]

In sum, the evidence reviewed in Part I makes a prima facie case against any facile claim of colorblindness, either cognitive or behavioral. Thus, when

---

105. The correlation is r=0.24. Squaring the r (r^2=5.7 percent) explains the percentage of variance, which reveals the extent to which differences in behavior can be attributed to a single variable (implicit bias). *Id*. at 24.

106. *See* Gregory Mitchell & Philip Tetlock, *Facts Do Matter: A Reply to Bagenstos*, 37 HOFSTRA L. REV. 737, 757–58 (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.").

107. If we assume that implicit measures have a test-retest reliability of approximately 0.56, and that the behavior has a reliability of 0.80, 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 that adjusts for error or "noise" in measurement of 0.41. This would then explain 16.7 percent of the variability in behavior. To focus again on behaviors within socially sensitive domains such as Black-White discrimination, a disattenuated correlation could rise from 0.24 to 0.36. The 0.56 test reliability comes from Nosek et al., *supra* note 28, at 274. The 0.80 behavior reliability is an estimate with no strong empirical basis, but it is the same number used by Greenwald et al., *supra* note 102, at 29, in their similar calculation. The formula for this new variable—the correlation corrected for attenuation—is $r_{xy}/\sqrt{(\text{Measure } x_{\text{reliability}} \text{ x } y_{\text{reliability}})}$. Notably, as the reliability of the measures decreases, the corrected correlation increases. Thus, an estimate of behavior's reliability of 0.80 is a conservative choice in calculating the disattenuated correlation.

108. *See* Robert P. Abelson, *A Variance Explanation Paradox: When a Little Is a Lot*, 97 PSYCHOL. BULL. 129, 132 (1985). Abelson used a simulation of a 0.320 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 0.220 hitter (a mediocre performance achieved by 221 of the 226 players with at least 375 plate appearances during the 2008 season). *See MLB Player Batting Stats – 2008*, ESPN.COM, http://www.espn.com/mlb/stats/batting/_/year/2008/minpa/375 (last visited Nov. 18, 2010).

109. 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.

we hear the stories of Juan, Skip, and Grace, we should be skeptical of any defense of strict colorblindness. Suppose that the supporting evidence accumulates into a widespread scientific consensus, especially regarding behavioral consequences in the real world. What impact should this consensus have on law and legal institutions?

II. LEGAL UPTAKE

A. Behavioral Realism

In answering such questions, our general approach is to 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 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."[110]

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.[111] How certain,

110. Kristin A. Lane et al., *Implicit Social Cognition and Law*, 3 ANN. REV. LAW & SOC. SCI. 427, 440 (2007). *See also* Linda Hamilton Krieger & Susan T. Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 CALIF. L. REV. 997, 1000 (2006) (finding that behavioral realism, understood as a prescriptive theory of judging, stands for the proposition that as judges develop substantive legal doctrines, they should guard against basing their analyses on inaccurate conceptions or irrelevant real-world phenomena).

111. 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 ENG. J. MED. 1784, 1785 (2005); Gina Kolata, *Mammograms Validated as Key in Cancer Fight*, N.Y. TIMES, Oct. 27, 2005, at A1.

then, must a behavioral realist be before calling the new view more "accurate"? Should we adopt, for instance, a "precautionary principle"?[112]

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 such "folk theory" is challenging.[113]

The third step, which demands an "accounting," is arguably the most complex. The force 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."[114] 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.

In fact, the only real normative commitment of behavioral realism is "against hypocrisy and self-deception."[115] 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 (i.e., 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."[116] That modern authority was a string cite to the psychological discoveries on the harm to Black children caused by segregation.[117]

Notwithstanding compelling new evidence, sometimes the status quo must prevail. But the countervailing reasons should be spelled out and publicly defended; otherwise, we risk hypocrisy and self-deception in our law. With

---

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

113. *See, e.g.*, Gary Blasi, *Advocacy Against the Stereotype: Lessons From Cognitive Social Psychology*, 49 UCLA L. REV. 1241, 1270–71 (2002) (discussing folk theories of discrimination).

114. *See* Lane et al., *supra* note 110, at 440–41.

115. Kang & Banaji, *supra* note 2, at 1065.

116. The relevant portion of the footnote reads: "Whatever may have been the extent of psychological knowledge at the time of *Plessy v. Ferguson*, this finding [of harm to Black children] 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." *Cf.* RICHARD KLUGER, SIMPLE JUSTICE 321 (1975) (describing Chief Justice Warren's addition of this footnote as an afterthought to repudiate *Plessy*'s stigma point). For a thoughtful analysis of science and civil rights, see ANGELO N. ANCHETA, SCIENTIFIC EVIDENCE AND EQUAL PROTECTION OF THE LAW (2006).

117. *See Brown*, 347 U.S. at 495 n.11 (citing, for example, K.B. Clark's doll studies).

behavioral realism, then, as our general approach, we examine how law and legal institutions can start seeing through colorblindness.

B. Four Quadrants of Legal Interventions

Even when readers are persuaded by the science, they are understandably anxious about what legal and policy prescriptions are thereby entailed. In particular, how are we supposed to react to the probabilistic nature of these cognitive tendencies? This raises a problem of specificity. For instance, we might have a general scientific understanding that implicit stereotypes cause differential employee evaluations at some given correlation. However, in an employment discrimination lawsuit, whether a particular applicant was treated worse by a particular reviewer is a different and much harder question.[118]

Another important distinction, which has been mostly ignored, is whether we are approaching the problem ex ante or ex post (a problem of time orientation). We generally enjoy greater flexibility to adopt ex ante interventions to prevent problems than to place legal liability or moral responsibility ex post. Cross-tabulating "specificity" against "time orientation" produces a Four-Quadrant model of possible legal interventions:[119]

		Time Orientation	
		Ex post	***Ex ante***
Specificity	***Specific***	I. *"Prejudice Polygraph"*	III. *Self-analysis*
	General	II. *Changing the Frame*	IV. *Prevention*

118. *See*, *e.g.*, David L. Faigman, *The Limits of Science in the Courtroom*, *in* BEYOND COMMON SENSE: PSYCHOLOGICAL SCIENCE IN THE COURTROOM 303 (Eugene Borgida & Susan T. Fiske eds., 2008); SHEILA JASANOFF, SCIENCE AT THE BAR: LAW, SCIENCE, AND TECHNOLOGY IN AMERICA 121 (1995) (discussing difficulties of showing general versus specific causation in toxic tort contexts). As we're using it, this general/specific distinction is related, but not identical, to the legislative/adjudicative fact distinction first offered by Kenneth Culp Davis. *See* Kenneth Culp Davis, *An Approach to Problems of Evidence in the Administrative Process*, 55 HARV. L. REV. 364, 402–04 (1942). Davis wrote that "[t]he rules of evidence for finding [legislative] facts which form the basis for creation of law and determination of policy should differ from the rules for finding [adjudicative] facts which concern only the parties to a particular case." *Id.* at 402. *Cf.* John Monahan & Laurens Walker, *Social Authority: Obtaining, Evaluating, and Establishing Social Science in Law*, 134 U. PA. L. REV. 477, 478, 488 (1986) (distinguishing "social authority" from "social facts").

119. This model draws on Jerry Kang, *The Missing Quadrants of Anti-discrimination: Going Beyond the "Prejudice Polygraph"*, J. SOC. ISSUES (forthcoming 2011).

1. "Prejudice Polygraph"?

Our cultural penchant for high-drama litigation naturally highlights Quadrant I, which involves specific facts in an ex post time orientation. The prototypical example is a lawyer waving some IAT score or fMRI neuroimage in front of a jury to show that a specific employer must have violated antidiscrimination law. In this example, a research instrument such as the IAT is being deployed as a sort of "Prejudice Polygraph."

This example belongs in Quadrant I because it raises a question about specific facts—indeed, specific in two senses. First, did this specific decisionmaker harbor implicit bias? Second, did the implicit bias affect the specific employment decision in question? The question is also ex post in that it is raised in a lawsuit requesting relief for past action.

We want to be crystal clear: No serious scientist has called for using instruments such as the IAT in this specific, ex post context. To the contrary, leading implicit bias scientists have testified directly against it for reasons outlined above.[120] In our view, the science as of 2010 simply doesn't support such an application.

But this just starts the conversation.

2. Changing the Frame

Quadrant II preserves the ex post time orientation but changes its focus to general facts. General facts are generalizations; they are probabilistic, based on aggregated observations. They transcend the relationship between the particular defendants and particular plaintiffs. We explore two contexts—social framework and structural litigation—in which general facts are either indirectly or directly relevant.

a. Social Framework Evidence

Social framework evidence "uses general conclusions from tested, reliable, and peer-reviewed social science research and applies them to the case."[121] When social scientists have gained a better understanding of general facts, especially when contrary to conventional wisdom, these findings can be shared in the form of expert testimony. The purpose is to supply general facts that inform

120. *See* Vedantam, *supra* note 50, at W14.

121. *See* Eugene Borgida & Susan T. Fiske, Introduction, BEYOND COMMON SENSE: PSYCHOLOGICAL SCIENCE IN THE COURTROOM, *supra* note 118, at xxxiii.

the jury and help it to reach its finding of specific facts.[122] In this way, general facts are deployed indirectly in the service of finding specific facts.

A prominent example is psychologist Susan Fiske's testimony in *Price Waterhouse v. Hopkins*.[123] That case involved Ann Hopkins, who was not promoted to partner, notwithstanding remarkable performance, in part because she was viewed as too masculine.[124] At trial, Fiske testified that, among other things, resentment and dissatisfaction could arise from a perceived lack of fit between the person's category (woman) and occupation (hard-charging manager).[125] Price Waterhouse dismissed such testimony as a "chain of intuitive hunches about 'unconscious' sexism."[126]

The trial court found liability. On appeal, the D.C. Circuit Court of Appeals affirmed and made clear that implicit biases mattered: "[U]nwitting or ingrained bias is no less injurious or worthy of eradication than blatant or calculated discrimination."[127] Furthermore, "the fact that some or all of the partners at Price Waterhouse may have been unaware of that motivation, even within themselves, neither alters the fact of its existence nor excuses it."[128] On certiorari, a plurality of the Supreme Court affirmed the finding that sex stereotyping played a part in Hopkins' evaluation.[129] Writing for the plurality, Justice Brennan was "tempted to say that Dr. Fiske's expert testimony was merely icing on Hopkins' cake" and that "no special training"[130] was necessary to discern the sex discrimination.

If general psychological discoveries about gender discrimination can be admitted as social framework evidence in *Price Waterhouse*, general findings about implicit bias and its undermining of colorblindness should be similarly

122. Expert testimony could be admitted regarding the general scientific evidence without allowing the expert to state any ultimate opinion on the specific causation at issue. This often happens with expert testimony on the unreliability of cross-racial identification. *See* David L. Faigman et al., *A Matter of Fit: The Law of Discrimination and the Science of Implicit Bias*, 59 HASTINGS L.J. 1389, 1393 (2008).

123. 490 U.S. 228 (1989). *See generally* Susan T. Fiske et al., *Social Science Research on Trial: Use of Sex Stereotyping Research in* Price Waterhouse v. Hopkins, 46 AM. PSYCHOLOGIST 1049 (1991).

124. *Price Waterhouse*, 490 U.S. at 235 (describing how Hopkins was told to "walk more femininely, talk more femininely, dress more femininely, wear make-up, have her hair styled, and wear jewelry").

125. *See* Fiske et al., *supra* note 123, at 1050.

126. *Id.* at 1053 (quoting Price Waterhouse's brief). The American Psychological Association took umbrage and filed an amicus brief to support the credibility of the methodology and literature Fiske used.

127. Hopkins v. Price Waterhouse, 825 F.2d 458, 469 (D.C. Cir. 1987).

128. *Id.*

129. *Price Waterhouse*, 490 U.S. at 255–58. The Court, however, reversed on the appropriate standard by which an employer would have to show that it would have made the same decision absent the discrimination. The courts below had required a "clear and convincing" standard; the plurality demanded only "preponderance of the evidence." *See id.* at 252–54. Thus, the case was reversed and remanded. *Id.* at 255.

130. *Id.* at 256. In this case, there was also plenty of evidence of explicit bias.

admissible.[131] The same evidentiary standards that admitted the former testimony should admit the latter.[132]

b. Structural Litigation

Structural reform litigation is another context that is ex post, yet involves general factual findings. For example, in *Chin v. Runnels*,[133] a habeas corpus petitioner challenged the San Francisco Superior Court's grand jury selection process for never having selected a foreperson of Chinese, Filipino, or Latino descent for thirty-six years. Given the deferential standard of review,[134] the federal court accepted the California appellate court's ruling that the government had successfully rebutted the prima facie case of discrimination and thus had not violated Chin's equal protection rights. However, in dicta, the federal court wrote extensively to explain why under a de novo standard of review there could have been a different result.[135] The court specifically cited a "growing body of social science [that] recognizes the pervasiveness of unconscious racial and ethnic stereotyping and group bias."[136]

---

131. *See, e.g.*, Faigman et al., *supra* note 122, at 1430–31 (arguing that social framework evidence regarding implicit bias should be admissible under Federal Rule of Evidence 702 and *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), as long as the expert did not attempt to opine regarding specific facts concerning a specific case). *Cf.* William T. Bielby, *Can I Get a Witness? Challenges of Using Expert Testimony on Cognitive Bias in Employment Discrimination Litigation*, 7 EMP. RTS. & EMP. POL'Y J. 377, 383 (2003) (arguing that "the social science scholarship on gender bias, stereotypes, and the structure and dynamics of gender inequality in organizations . . . has substantial external validity" and should be admissible); *id.* at 389 ("[I]t seems appropriate to allow introduction of this evidence by qualified experts, leaving to the jury the question of whether the plaintiff has proved the existence of stereotypes in a given case.").

132. In federal courts, judges act as the gatekeepers for such scientific evidence, in accordance with Federal Rule of Evidence 702. Under Rule 702, an expert may testify in the form of an opinion or otherwise "if (1) the testimony is based upon sufficient facts or data; (2) the testimony is the product of reliable principles and methods; and (3) the witness has applied the principles and methods reliably to the facts of the case." FED. R. EVID. 702. This rule essentially codifies the Supreme Court's *Daubert-G.E.-Kumho* trilogy of cases, which requires a flexible determination about the reliability of science. *See Daubert*, 509 U.S. 579; Gen. Elec. Co. v. Joiner, 522 U.S. 136 (1997); Kumho Tire Co., Ltd. v. Carmichael, 526 U.S. 137 (1999).

133. 343 F. Supp. 2d 891 (N.D. Cal. 2004). For law review discussion of *Chin v. Runnels*, see Darren Seiji Teshima, *A "Hardy Handshake Sort of Guy": The Model Minority and Implicit Bias About Asian Americans in* Chin v. Runnels, 11 ASIAN PAC. AM. L.J. 122 (2006).

134. In this habeas context, the court explained that it "must find that the state court's decision was 'objectively unreasonable.'" *Chin*, 343 F. Supp. 2d at 901 (quoting Lockyer v. Andrade, 538 U.S. 63, 75 (2003)).

135. *Id.* at 903.

136. *Id.* at 906. Many of the citations to the academic literature were to early, important Critical Race Theory literature; other references were made to the more social cognitive literature.

Another example is *Farrakhan v. Gregoire*,[137] which involved a section 2 Voting Rights Act (VRA)[138] challenge to Washington state's felon disenfranchisement statute.[139] Among the necessary (but not sufficient) conditions for plaintiffs to prevail was a factual finding that the Washington criminal justice system had engaged in racial discrimination. Again, this is a general fact finding, not a specific one focused on particular cops, prosecutors, juries, or judges.

Both parties cross-motioned for summary judgment. It is breathtaking for a branch of the federal government to call an entire state's criminal justice apparatus biased, but that's more or less what the trial court said: The district court found "compelling evidence of racial discrimination and bias in Washington's criminal justice system."[140]

According to the district court, this finding was not based "solely on [racial disparity] statistics";[141] it was also based on expert testimony that fundamentally changed the frame by unpacking structural, institutional, and implicit social cognitive mechanisms for discrimination. The court discussed, for example, the expert report from sociology professor Robert Crutchfield, who posited two possible explanations for racial disparities: "(1) discriminatory actions of criminal justice decision makers (either intentional or unconscious); and (2) structural or institutional causes (ways of doing business, such as decision rules that are theoretically race-neutral, but are not race-neutral in practice)."[142] Among various factors, the court noted Crutchfield's identification of "implicit biases."[143] In a footnote, the court also referred to other reports—which included a "draft law review article discussing the concept of implicit bias" written by Anthony Greenwald—that helped "bolster the Court's conclusion."[144]

137. No. CV-96-076-RHW, 2006 U.S. Dist. LEXIS 45987, at *3 (E.D. Wash. July 7, 2006).

138. *See* 42 U.S.C. § 1973(a) (2006) ("No voting qualification or prerequisite to voting or standard, practice, or procedure shall be imposed or applied by any State or political subdivision in a manner which results in a denial or abridgement of the right of any citizen of the United States to vote on account of race or color . . . .").

139. *See* WASH. CONST. art. VI, § 3 ("All persons convicted of infamous crime unless restored to their civil rights . . . are excluded from the elective franchise.").

140. *Farrakhan*, 2006 U.S. Dist. LEXIS 45987 at *17.

141. *Id.*

142. *Id.* at *14. The report submitted by sociology professor Katherine Beckett was also discussed. *See id.* at *18 n.6.

143. *Id.* at *15.

144. *Id.* at *18 n.6. Greenwald submitted a draft of his contribution to the Behavioral Realism symposium published in the *California Law Review*. *See* Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CALIF. L. REV. 945 (2006).

We do not want to overread the importance of this case.[145] Most important, notwithstanding the damning finding, the district court granted summary judgment to the state under a totality of the circumstances test.[146] And although that judgment was initially reversed on appeal, the district court's judgment was ultimately affirmed by the Ninth Circuit Court of Appeals sitting en banc.[147] Nonetheless, *Farrakhan* demonstrates how the reality of contemporaneous discrimination, described through social science (ranging from sociology to psychology), can assist courts in finding general facts that could be legally consequential.

Both *Chin* and *Farrakhan* are examples of "structural litigation," by which we mean litigation that critiques some structural feature of a system, such as prison conditions, voting procedures, or segregated education.[148] In such litigation, the legal problem is understood as the aggregation of myriad individual transactions. The general factual finding about the discrimination in California's grand jury selection or Washington's criminal justice system transcends the experience of plaintiffs Chin or Farrakhan. This conceptualization, in turn, permits the use of averages, probabilities, and general facts.[149]

Our last example is historical: *Brown v. Board of Education*.[150] There, the Court wrote: "Whatever may have been the extent of psychological knowledge

145. As the Ninth Circuit pointed out, the State of Washington conducted a risky litigation strategy by not trying to produce a genuine issue of material fact through contradictory expert testimony. *See* Farrakhan v. Gregoire, 590 F.3d 989, 1003 (9th Cir. 2010). Instead, it conceded that no genuine issue existed but that it (not the plaintiffs) should prevail as a matter of law.

146. *Farrakhan*, 2006 U.S. Dist. LEXIS 45987 at *29.

147. On appeal, a three-judge panel of the Ninth Circuit Court of Appeals reversed and granted summary judgment to the plaintiffs. *Farrakhan*, 590 F.3d at 1016, *reh'g granted*, 603 F.3d 1072 (9th Cir. 2010) (en banc). The court of appeals referred repeatedly to the expert evidence but made no specific mention of implicit bias. *See id.* at 994–95, 1009–12 (discussing the work of Robert Crutchfield and Katherine Beckett). Subsequently, the entire Circuit ordered the case to be reheard en banc, where the district court's grant of summary judgment to the State of Washington was affirmed. Farrakhan v. Gregoire, No. 06-35669, 2010 WL 4054429 (9th Cir. Oct. 7, 2010) (en banc) (per curiam) (holding as a matter of law that "plaintiffs bringing a section 2 VRA challenge to a felon disenfranchisement law based on the operation of a state's criminal justice system must at least show that the criminal justice system is infected by *intentional* discrimination or that the felon disenfranchisement law was enacted with such intent") (emphasis in original).

148. For a description of the terms public law litigation, structural reform litigation, and institutional reform litigation, see Susan P. Sturm, *A Normative Theory of Public Law Remedies*, 79 GEO. L.J. 1355, 1357 n.1, 1385–87 (1991). *See also* Margo Schlanger, *Beyond the Hero Judge: Institutional Reform Litigation as Litigation*, 97 MICH. L. REV. 1994, 1995 (1999).

149. Similar reasoning may apply to some class action cases. *See* Bielby, *supra* note 131, at 395 ("[I]n a class action case, the social science expert is not asserting that *every single* personnel decision adversely affected women compared to similarly situated men . . . . Nor in the typical case is the expert asserting that *any specific personnel action* was, with certainty, adversely affected by gender bias. Instead, he or she is more likely to be claiming that . . . it is more probable than not that personnel decisions, taken as a whole, are likely to have favored men over women.") (emphasis added).

150. 347 U.S. 483 (1954).

at the time of *Plessy v. Ferguson*, this finding [of harm to Black children] is amply supported by modern authority."[151] That "modern authority" referred to psychological studies that found general facts.[152] These were not specific facts found about the stigma suffered by the specific plaintiff children as determined via clinical diagnosis. Yet, these general facts (regarding stigma) in an ex post setting (constitutional litigation challenging segregated education) helped craft a unanimous opinion that signaled the demise of de jure school segregation.[153]

These examples of social framework evidence and structural reform litigation demonstrate that general facts can be relevant in certain cases, sometimes indirectly as a way to help the finding of specific facts, and sometimes more directly, as an element of the cause of action. In such cases, the scientific dismantling of colorblindness matters—even in the context of ex post accountability.

3. Self-Analysis

Quadrant III is the least familiar of the four: specific facts, but ex ante. To grasp an example, let's recall *Skip's Encounter*. There is some evidence that particular training regimens can decrease shooter bias.[154] So, suppose that a socially responsible police chief investigates whether her officers have shooter bias. Finding that they do, she adopts new training that decreases that bias. After a shooting in the field of an African American youth, the plaintiff in a civil suit requests the shooter bias scores of the accused police officer.

In federal court, Federal Rule of Civil Procedure 26(b)(1) permits discovery of "nonprivileged matter that is relevant to any party's claim or defense."[155] The requested material need not itself be admissible as long as it is "reasonably calculated to lead to the discovery of admissible evidence."[156] Under this generous standard, the shooter bias self-measurements may well be discoverable. But disclosing this information would create perverse incentives to avoid discovering one's biases in the first place—in which case the countermeasures to decrease shooter bias would have never been adopted.

To avoid this fate, we could adopt some sort of self-analysis privilege. Similar privileges have been recognized in various states. For example, some states

151. *Id.* at 494.
152. *See id.* at 494 n.11 (citing, for example, K.B. Clark's doll studies).
153. *See supra* Part II.A.
154. E. Ashby Plant et al., *Eliminating Automatic Racial Bias: Making Race Non-Diagnostic for Responses to Criminal Suspects*, 41 J. EXPERIMENTAL SOC. PSYCHOL. 141, 153 (2005).
155. FED. R. CIV. P. 26(b)(1).
156. *Id.*

have recognized a privilege for medical committee reports and self-evaluations after a medical accident.[157] Federal Rule of Evidence 407 also recognizes a privilege that covers "subsequent remedial measures."[158] Deana Pollard has specifically called for a similar evidentiary privilege covering self-discovery of implicit biases.[159]

Since leading scientists have rejected using the IAT and similar instruments for Quadrant I application, keeping specific scores about specific people out of the courtroom would not much undermine the goal of truth seeking.[160] Again, this brief discussion cannot and was not intended to weigh definitively the costs and benefits of creating this evidentiary privilege. Instead, our point is merely to show a concrete example of a potential legal intervention in Quadrant III.

4. Prevention

Finally, consider Quadrant IV—general facts and ex ante time orientation, conceptually most distant from any "Prejudice Polygraph." The paradigmatic context for this quadrant is legislation or administrative rulemaking, not a jury trial adjudication. As for specificity, when a legislature enacts legislation (e.g., insisting that library computers have software filters) to prevent some problem (e.g., minors being exposed to indecency), how does it know that the problem exists in the first place, the magnitude of the problem, and its likelihood? How does it know that its proposed legislation will help in an efficient way, without unintended consequences? These questions all necessarily turn on facts, but they are not the specific facts of any particular isolated case or adjudication. Instead, they are more general facts, similar to the ones that scientists discover. As for time orientation, when the statute is enacted, the point is not to hold someone liable for previous actions; instead, it is to create a law that changes prospectively some collective behavior of interest.

In this quadrant lie the possibilities of revising the law or behavior in light of better scientific understanding of our general cognitive tendencies—in short, taking sensible steps toward prevention. For instance, if the goal is to be

157. *See* James F. Flanagan, *Rejecting a General Privilege for Self Critical Analyses*, 51 GEO. WASH. L. REV. 551, 557 n.136 (1983).

158. After a product has hurt someone, a manufacturer might want to improve that product's design. But if such remedy will be exploited by a trial lawyer as tacit admission of the product's defect, a manufacturer might think twice. To decrease any such disincentive, the law prevents the evidence of subsequent remedial measures from reaching the jury to prove negligence or defect.

159. *See* Deana A. Pollard, *Unconscious Bias and Self-Critical Analysis: The Case for a Qualified Evidentiary Equal Employment Opportunity Privilege*, 74 WASH. L. REV. 913, 915–16 (1999).

160. Even if implicit bias scores remain stable over time, it would be very difficult to know whether implicit bias was a partial cause in any particular shooting.

behaviorally colorblind, what kinds of laws, practices, procedures, and education might help? Consider just a few examples.

a. Debiasing the Courtroom

Sometimes we seek behavioral colorblindness in the short term—e.g., how judges[161] and juries think through a criminal matter in which race is in the air.[162] Could a particular program of juror education (while waiting to be called for jury duty), specific debiasing jury instructions, or even a courtroom setting promote such colorblindness?[163] According to Gary Blasi, there is good reason to think that the answer is yes.[164] As explained above, being aware of potential biases, being motivated to check those biases, and being accountable to a superior (as a jury feels toward a judge) should have some effect on the translation of bias to behavior.

Federal district court judge Mark Bennett extensively discusses the problem of implicit bias with potential jurors during voir dire,[165] and in both civil and criminal cases, he instructs on implicit bias as part of his complete set of jury instructions before opening statements.[166] He also requires jurors to sign a

---

161. *See, e.g.*, Irene V. Blair et al., *The Influence of Afrocentric Facial Features in Criminal Sentencing*, 15 PSYCHOL. SCI. 674, 677 (2004) (finding no disparate sentencing on the basis of race in a Florida data set but finding that within each racial category, White or Black, those individuals with more Afrocentric facial features received harsher sentences).

162. *See, e.g.*, Samuel R. Sommers & Phoebe C. Ellsworth, *White Juror Bias: An Investigation of Prejudice Against Black Defendants in the American Courtroom*, 7 PSYCHOL. PUB. POL'Y & L. 201, 210 (2001) (showing through experiment that White jurors are more likely to demonstrate racial prejudice in cases without salient racial issues). An interracial locker room fight narrative produced racially biased results (with mock jurors convicting Black defendants more often than White defendants under identical facts). A single additional fact that made race more explicitly salient—that the defendant had been subject to racial remarks from teammates—removed the racial bias in juror results. For summary in the law reviews, see Blasi, *supra* note 113, at 1246–47.

163. *See, e.g.*, Justin D. Levinson, *Forgotten Racial Equality: Implicit Bias, Decisionmaking, and Misremembering*, 57 DUKE L.J. 345, 413–14 (2007) (discussing various debiasing techniques, including confronting jurors with biases); *id.* at 415–17 (discussing juror diversity training).

164. *See* Blasi, *supra* note 113, at 1276–77 (summarizing relevant psychological research). Blasi emphasizes that the problem of jury bias should be addressed both explicitly and implicitly. *See id.* at 1277–79. As for the explicit strategy, he concludes "that there is good reason explicitly to instruct juries in every case, stereotype-salient or not, about the specific potential stereotypes at work in the case." *Id.* at 1277.

165. *See* Mark W. Bennett, *Unraveling the Gordian Knot of Implicit Bias in Jury Selection: The Problems of Judge-Dominated Voir Dire, the Failed Promise of* Batson*, and Proposed Solutions*, 4 HARV. L & POL'Y REV. 149, 169 (2010).

166. Judge Bennett gives the following instruction:

> As we discussed in jury selection, everyone, including me, has feelings, assumptions, perceptions, fears, and stereotypes, that is, "implicit biases," that we may not be aware of. These hidden thoughts can impact what we see and hear, how we remember what we see and hear, and how we make important decisions. Because you are making very important decisions in this case, I strongly encourage you to evaluate the evidence carefully and to resist jumping to conclusions based on

certification of nondiscrimination in criminal cases.[167] Jurors are already instructed to be fair and square; if research demonstrates that such a specifically tailored instruction or an accountability-promoting certification can counter implicit bias even better, then such general facts should drive their broader ex ante adoption.

b. Debiasing the Classroom

Some Quadrant IV prevention strategies are longer-term. For example, Jerry Kang and Mahzarin Banaji have raised the possibility of hiring certain individuals in part because they function as "debiasing agents."[168] The scientific premise for such an idea is that exposure to countertypical exemplars appears to decrease implicit bias.

Three studies provide intriguing evidence. First, imagining a countertypical female leader (as opposed to a Caribbean vacation) for a few minutes in a mental imagery exercise reduced implicit gender stereotyping.[169] Second, exposure to positive exemplars of the disfavored social category (pictures of Tiger Woods[170] and Martin Luther King) decreased implicit bias against that category significantly in magnitude and over a twenty-four-hour period.[171] 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

> stereotypes, generalizations, gut feelings, or implicit biases. The law demands that you return a just verdict, based solely on the evidence, your individual evaluation of that evidence, your reason and common sense, and these instructions. Our system of justice is counting on you to render a fair decision based on the evidence, not on biases.

Email From the Hon. Mark W. Bennett, U.S. District Court Judge, Northern District of Iowa to Jerry Kang (Nov. 3, 2010) (on file with author).

167. *Id.* (duty during deliberations instruction) (on file with author) ("Fifth, in your consideration of whether the defendant is not guilty or guilty of the offense charged against him, you must not consider his race, color, religious beliefs, national origin, or sex. You are not to return a verdict for or against the defendant on any charge unless you would return the same verdict on that charge without regard to the defendant's race, color, religious beliefs, national origin, or sex. To emphasize the importance of this consideration, the verdict form contains a certification statement. Each of you should carefully read the statement, then sign your name in the appropriate place in the signature block, if the statement accurately reflects the manner in which each of you reached your decision.").

168. Kang & Banaji, *supra* note 2, 1109–15.

169. Irene V. Blair, Jennifer E. Ma & Alison P. Lenton, *Imagining Stereotypes Away: The Moderation of Implicit Stereotypes Through Mental Imagery*, 81 J. PERSONALITY & SOC. PSYCHOL. 828, 831–32 (2001).

170. Given Tiger Woods's recent bad press for sexual indiscretion, it's not clear whether he would function as a debiasing agent in the current media atmosphere. In the original experiment, O.J. Simpson was offered as a schema-consistent (not a debiasing) exemplar. Here's where cultural studies meet social cognition. *See* Kang, *supra* note 14, at 1582.

171. Nilanjana Dasgupta & Anthony G. Greenwald, *On the Malleability of Automatic Attitudes: Combating Automatic Prejudice With Images of Admired and Disliked Individuals*, 81 J. PERSONALITY & SOC. PSYCHOL. 800, 802–05 (2001); Mitchell et al., *supra* note 53.

zero.[172] By contrast, a control group of women who attended a coed university for one year had its average group implicit bias increase.[173] What explained these movements? After examining various university environmental variables, such as coursework and extracurriculars, Nilanjana Dasgupta and Shaki Asgari determined that the mechanism was exposure to female professors and administrators.[174]

Such science could lead an institution, such as a public law school, to break a tie and hire a racial minority applicant (over a comparably qualified White applicant)—not as a "role model" but as a debiasing agent.[175] Recall *Grace's Hire*. Such a decision could draw an equal protection challenge, reviewed under strict scrutiny.[176] First, the end sought by the state actor has to be "compelling." Second, the means deployed—in this case exposure to a countertypical exemplar—must be "narrowly tailored" to achieve that compelling interest.

The end sought by this intervention is not to inflate minority student self-esteem via role modeling; to redistribute goodies between racial groups or genders; or to correct hard-to-measure general societal discrimination. The Supreme Court has rejected all three such goals as not quite compelling.[177] Rather, the point of debiasing is to directly counter racial bias that has been found to predict discrimination. And fighting racial discrimination has always been recognized as a compelling interest.[178]

When determining whether the means are narrowly tailored, courts typically examine whether the technique deployed is underinclusive, overinclusive, and whether some alternative that is not facially race-conscious could have solved the problem just as well. Here, the particulars of the debiasing intervention and its efficaciousness will determine the final analysis. We simply observe that courts have accepted science in other contexts where the evidence

172. *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. EXPERIMENTAL SOC. PSYCHOL. 642, 649–54 (2004).

173. *Id.*

174. *Id.*

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

176. Of course, in this hypothetical, the university is a state actor.

177. *See*, *e.g.*, Wygant v. Jackson Bd. of Educ., 476 U.S. 267 (1986) (plurality opinion); City of Richmond v. J.A. Croson Co., 488 U.S. 469, 497–98 (1989) (plurality opinion) (citing and summarizing *Wygant's* view of role models); Taxman v. Piscataway Twp. Bd. of Educ., 91 F.3d 1547 (3d Cir. 1996) (en banc) (rejecting role model justification). *Wygant*, 476 U.S. at 274 (plurality opinion) ("This Court never has held that societal discrimination alone is sufficient to justify a racial classification.").

178. We concede that fighting racial discrimination caused by implicit bias might be begging the question, since for some, such discrimination isn't and shouldn't be legally cognizable. *See* Samuel R. Bagenstos, *Implicit Bias, "Science," and Antidiscrimination Law*, 1 HARV. L. & POL'Y REV. 477, 488 (2007) (discussing the fact that many people adopt an irrational-animus theory of antidiscrimination law).

of narrow tailoring is arguably no better. Consider specifically the doll studies in *Brown* (striking down "separate but equal") and the educational diversity studies in *Grutter* (upholding the admissions policy at Michigan Law School).[179] Of course, these opinions could be interpreted as more political compromise than earnest application of new science.

We do not want to be misread. The goal in discussing these examples is not to persuade anyone of the complex merits of any particular policy. Instead, it is to outline examples and opportunities that may unfold under Quadrant IV, which focuses on prevention—a goal that could trigger broad support across the political spectrum.

* * *

In sum, the empirical assertion of colorblindness is being dismantled by the findings in implicit social cognition. As behavioral realists, we believe that this new recognition should have social and legal implications. This does not mean, however, that we are calling for "Prejudice Polygraphs" or mind readers of any sort. That said, there are plenty of domains, such as "Prevention," in which a more scientifically nuanced analysis regarding colorblindness and color consciousness can inform and influence the path of law. Judges are already studying what might be done.[180]

III. OBJECTIONS

Of course, everything we have written may be politically naive. We have described the science with little mention of how cultural, social, and political forces can influence not only which questions are asked, but which answers are accepted.[181] We have assumed that people actually care about implicit-bias-actuated behaviors; to the contrary, people may care only about behaviors caused by explicit biases,[182] and we may need entirely new "normative underpinnings for

179. *See* Kang & Banaji, *supra* note 2, at 1112.

180. For instance, the National Center for State Courts has a working group on implicit bias and has helped produce a primer on the subject for judges. *See* NAT'L CTR. FOR STATE COURTS, http://www.ncsconline.org/D_Research/ref/implicit.html (last visited Nov. 10, 2010).

181. *See, e.g.*, David S. Caudill, *Ethnography and the Idealized Accounts of Science in Law*, 39 SAN DIEGO L. REV. 269, 273–74 (2002) (explaining how *Daubert* and its codification idealize science when "each of the core features of science also has an anchor in social structures"); *see also id.* at 279 (Diagram II) (providing a map of both external and internal influences on peer review and general acceptance of science). *See generally* JASANOFF, *supra* note 118, at 8 (observing the mutually constitutive relationship between law and science).

182. *See, e.g.*, *supra* note 12. No doubt, many would like to read the disparate treatment strand of Title VII to cover only "intentional" discrimination—that is, behavior caused by explicit biases that are

the antidiscrimination law."[183] We have advocated for a behavioral realism in which legal actors dutifully take account of new science, as if somehow the science overdetermines the specific policy response. Perhaps, none of this will come to be. Instead, the implicit social cognitive challenge to colorblindness could get dismissed along three lines of objections: "junk science" backlash, "hardwired" resignation, and "rational" justification.

A. "Junk Science" Backlash

Statistics lie. Lawyers speak out of both sides of their mouth. Even scientists fudge data. Maybe all this so-called "science" is nothing but "junk." So argues the "junk science" backlash, which characterizes the implicit bias research as politically motivated by junk scientists and their lawyer accomplices who manipulate data, misinterpret results,[184] and exaggerate findings[185] in order to

introspectively recognized and endorsed as such by the actor. *See*, *e.g.*, Krieger & Fiske, *supra* note 110, at 1035–36 (explaining the diffusion of the "honest belief rule" in Title VII law, which states that a "plaintiff could not prevail if the decision maker 'honestly believed in the non-discriminatory reasons it offered, even if the reasons are foolish or trivial or even baseless'") (footnote omitted).

183. Samuel R. Bagenstos, *The Structural Turn and the Limits of Antidiscrimination Law*, 94 CALIF. L. REV. 1, 3 (2006).

184. For example, Gregory Mitchell and Philip Tetlock suggest that the work of the "implicit prejudice scholars" (we prefer the less-loaded term implicit bias or implicit social cognition) shouldn't be called science. *See* Mitchell & Tetlock, *supra* note 86, 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 is 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 ("[T]he focus has shifted with prevailing intellectual *fashions* from psychodynamic theories to social-identity theories . . . .") (emphasis added); *id.* at 1062 ("by *now-out-of-fashion* psychodynamic theories") (emphasis added).

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 [i.e. Behavioral Law & Economics] is nothing more than a mess of overgeneralizations about how people exhibit this or that bias or anomaly under largely unspecified conditions."). In short, in 2002, Mitchell viewed Behavioral Law & Economics as too complex. *See id.* at 119–22.

Seven years later, however, Mitchell criticized Behavioral Law & Economics for being too simplistic. *See* Gregory Mitchell, *Second Thoughts*, 40 MCGEORGE L. REV. 687, 709 (2009) ("[C]urrent, popular models of judgment and decision-making within antidiscrimination theory and behavioral-law-and-economics theory portray humans in *too simplistic a light*; these models fail to give sufficient weight to the impact of second thoughts on first thoughts.") (emphasis added).

185. Regarding the implicit bias research, Mitchell and Tetlock lament: the "superficial and selective" reference to relevant factors, *supra* note 86, at 1108; the refusal to provide "full disclosure," *id.* at 1091; and the deployment of "specious but seductive" arguments, *id.* at 1100, that are revealed by "pulling back the curtain," *id.* at 1030.

snooker society into politically correct[186] 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.

It's hard to respond to junk science allegations. On the one hand, we could proffer still more evidence and engage still more on the scientific merits. On the other hand, such engagement is naive because junk science backlash trades on a double standard. The critics demand scientific rigor that is practically unachievable while not requiring the same of factual assumptions undergirding the status quo.

Gregory Mitchell and Philip Tetlock, the two most prominent critics of implicit bias work, deploy a double standard on when and how much scientific certainty is necessary. Throughout their paper warning about the "Perils of Mindreading,"[187] they are extraordinarily demanding on the implicit bias science. Almost no epistemic limitation goes unraised. They challenge each validity, methodology, definition, and dataset (probing for outliers). They do not, however, deploy the same deliberate skepticism to the studies whose conclusions they prefer. Each and every study that supposedly debunks the implicit bias "orthodoxy" is cited as if mere publication guaranteed a study's flawless methodology.

Here's one telling example regarding "stereotype threat." Near the end of their one-hundred-page article, they accuse implicit bias scholars of fomenting racial discord. Specifically, they warn that implicit bias researchers are "convincing Blacks that they are held in contempt, thereby inducing 'stereotype threat.'"[188] Earlier, they suggest that the IAT effect itself might be explained by stereotype threat, that "fear of being labeled a bigot" is what drives the results.[189] To support these fundamentally empirical claims, they cite a single study.[190]

Mitchell has complained about exaggerations before. For example, he suggests that behavioral economics exaggerates 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 (2002).

186. Mitchell and Tetlock complain that implicit bias scholars are driven by "moral certitude," Mitchell & Tetlock, *supra* note 86, at 1029, come from an "ideologically-skewed field," *id.* at 1064, and are 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. *Id.* at 1066, 1117–18.

187. *Id.*

188. *Id.* at 1119.

189. *Id.* at 1093.

190. *See id.* at 1079 n.184 (citing Cynthia M. Frantz et al., *A Threat in the Computer: The Race Implicit Association Test as a Stereotype Threat Experience*, 30 PERSONALITY & SOC. PSYCHOL. BULL. 1611 (2004)).

In making this argument, Mitchell and Tetlock have publicly embraced the science of stereotype threat, including its real-world predictive and ecological validity. Indeed, they warn that this "has the potential to create a self-fulfilling prophecy."[191] But what's puzzling is that stereotype threat is also an implicit social cognition. And no one has made the case that stereotype threat science has better predictive validity than implicit bias science. What, then, is the scientific basis for endorsing the stereotype threat findings while simultaneously dismissing the implicit bias findings? If Mitchell and Tetlock take stereotype threat that seriously, they should call for revamping how schools use standardized testing—something they have never publicly called for, and Amy Wax (a fellow skeptic of implicit bias) has sharply criticized.[192]

Worse, some forms of backlash aren't entirely interested in the merits. 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, the tobacco industry generated its own counter-science contesting the causation evidence.[193] It complained that the causation science was junk, unproven and insufficiently rigorous. In short, it manufactured scientific doubt, a deliberate strategy it had deployed for decades.[194] The tobacco industry even impugned the EPA's scientific integrity, calling its research "another step in a long process characterized by a preference for political correctness over sound science."[195] 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.

We want to be clear: We are *not* suggesting that any call for skepticism that challenges our particular scientific understanding, based on the best evidence as of 2010, 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 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.[196] Earlier discussions about reliability and validity reflected substantive

191. *Id.* at 1080.

192. *See* Amy L. Wax, *Stereotype Threat: A Case of Overclaim Syndrome?* (Univ. of Penn. Law School, Public Law Research Paper No. 08-14, 2008), *available at* http://ssrn.com/abstract=1123499.

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

194. *Id.* at 67.

195. *Id.* at 66.

196. Critiques from academic psychologists encouraged 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*

engagements in precisely this vein—both respectful and inviting on-the-merits skepticism. That said, backlash scholarship smells different.[197]

It somehow insists that without perfect knowledge about this "thing" called implicit bias, we must be prohibited from responding whatsoever. At the same time, backlash scholarship never requires perfect knowledge of the assumptions that underlie the status quo. The differential epistemological standards are never justified, nor is the absolute standard of "do nothing unless we know (almost) all." We think to the contrary.

Consider an example from the biomedical sciences. Chemotherapy agents Tarceva and Iressa have shown great efficacy against lung cancer despite the open acknowledgement that researchers don't understand their underlying mechanisms.[198] That scientists have only a general sense of how Tarceva and Iressa work (by inhibiting an epidermal growth factor receptor (EGFR)), does not, however, make the drug worthless to cancer patients. Indeed, initial clinical trials showed a "promising" response rate of "12% to 19%,"[199] in which "patients who did respond often had dramatic, rapid, and sustained improvement."[200] Subsequent research uncovered "inherently enriched" populations who would most likely benefit from the drug.[201] People in these groups had responses several

Variables and Construct Validity, 31 PERSONALITY & SOC. PSYCHOL. BULL. 166 (2005); Greenwald et al., *supra* note 29; *see also supra* note 60.

197. *See generally* Thomas O. McGarity, *Defending Clean Science From Dirty Attacks by Special Interests*, *in* RESCUING SCIENCE FROM POLITICS: REGULATION AND THE DISTORTION OF SCIENTIFIC RESEARCH 24, 24–39 (Wendy Wagner & Rena Steinzor eds., 2006) (cataloging the various techniques employed by risk-producing industries to deny causation). According to McGarity, these techniques include "attack science," "deconstruction through reanalysis," and going after individual faults in individual studies in a "corpuscular approach." *Id.*

198. Genentech, who markets the drug Tarceva in the United States, describes the "*Proposed* Mechanism of Action" (emphasis added) on its website, concluding that "[t]he clinical anti-tumor action of erlotinib is not fully characterized." *Tarceva*, GENENTECH, http://www.gene.com/gene/products/information/ oncology/tarceva (last visited Nov. 11, 2010). AstraZeneca (who owns the trademark of IRESSA, generic name gefitinib) describes in the "IRESSA Full Prescribing Information" that "[t]he mechanism of the clinical antitumor action of gefitinib is not fully characterized." ASTRAZENECA, IRESSA 2 (2005), *available at* http://www1.astrazeneca-us.com/pi/iressa.pdf. Although these are similar agents, given data that Iressa was not as effective in general populations as originally thought, the Federal Drug Administration has limited its use only to patients currently taking it who have shown a response. Andrew Pollack, *F.D.A. Restricts Access to Cancer Drug, Citing Ineffectiveness*, N.Y. TIMES, June 18, 2005, at C2.

199. Geoffrey R. Oxnard & Vincent A. Miller, *Use of Erlotinib or Gefitinib as Initial Therapy in Advanced NSCLC*, 24 ONCOLOGY 392, 392 (2010).

200. *Id.* at 393.

201. These groups include people of Asian descent, women, people with no (or a light) smoking history, and people with mutations in the EGFR gene. *See* Masahiro Fukuoka et al., *Multi-Institutional Randomized Phase II Trial of Gefitinib for Previously Treated Patients With Advanced Non-Small-Cell Lung Cancer*, 21 J. CLINICAL ONCOLOGY 2237, 2239, 2242 (2003); Mark G. Kris et al., *Efficacy of Gefitinib, an Inhibitor of the Epidermal Growth Factor Receptor Tyrosine Kinase, in Symptomatic Patients With Non-Small-Cell Lung Cancer: A Randomized Trial*, 290 JAMA 2149, 2152–53 (2003); Nick Thatcher et al., *Gefitinib Plus Best Supportive Care in Previously Treated Patients With Refractory Advanced Non-Small-Cell Lung*

orders of magnitude greater than in the original trials.[202] The basic point here is that science rarely provides perfect information, and demanding total understanding of every possible mechanism or boundary condition of a process will both guarantee inaction and stunt future research. And, to repeat, this stringent epistemological requirement is rarely required of the laws and policies embedded in the status quo.

In the end, what we can offer is a suggestion on how to better distinguish between legitimate (even if heated) disagreements and backlash.[203] First, whenever possible, questions of scientific quality should be decoupled from allegations of venality, fraud, or politicization. Second, we should respect the relative institutional competencies of the scientific versus legal communities. Journals that are peer-reviewed by scientific experts are the proper fora for determining whether something is good "science."[204] We see no reason why the operation of scientific orthodoxy would not produce a consensus over the long run.[205] It did so with global warming—notwithstanding "Climategate."[206] By

---

Cancer: Results From a Randomised, Placebo-Controlled, Multicentre Study (Iressa Survival Evaluation in Lung Cancer), 366 LANCET 1527, 1531 (2005).

202. Indeed, one study in Japan had a 30 percent response rate. Seiji Niho et al., *First-Line Single Agent Treatment With Gefitinib in Patients With Advanced Non-Small-Cell Lung Cancer: A Phase II Study*, 24 J. CLINICAL ONCOLOGY 64, 65 (2006). A study in Taiwan had an "impressive" 51 percent response rate. The forty-three patients in that study with the EGFR mutation had an 84 percent response rate, which is several orders of magnitude higher than the response rate in the general population. Chih-Hsin Yang et al., *Specific EGFR Mutations Predict Treatment Outcome of Stage IIIB/IV Patients With Chemotherapy-Naive Non-Small-Cell Lung Cancer Receiving First-Line Gefitinib Monotherapy*, 26 J. CLINICAL ONCOLOGY 2745, 2747–48 (2008). Other studies show similarly impressive results in patients with the EGFR mutation. In two Japanese studies, 75 percent of such patients responded to the drug, *see* Haruka Asahina et al., *A Phase II Trial of Gefitinib as First-Line Therapy for Advanced Non-Small-Cell Lung Cancer With Epidermal Growth Factor Receptor Mutations*, 95 BRIT. J. CANCER 998, 1000, 1002 (2006); Akira Inoue et al., *Prospective Phase II Study of Gefitinib for Chemotherapy-Naive Patients With Advanced Non-Small-Cell Lung Cancer With Epidermal Growth Factor Receptor Gene Mutations*, 24 J. CLINICAL ONCOLOGY 3340, 3342 (2006). In a Spanish study, 71 percent of patients responded to the drug. Rafael Rosell et al., *Screening for Epidermal Growth Factor Receptor Mutations in Lung Cancer*, 361 NEW ENG. J. MED. 958, 961–63 (2009).

203. Even in legitimate disagreements, self-interested reasoning surely abounds. 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). *See also* Sung Hui Kim, *The Banality of Fraud: Re-Situating the Inside Counsel as Gatekeeper*, 74 FORDHAM L. REV. 983, 1029–34 (2005) (discussing motivated reasoning).

204. We recognize that peer-review has its own various limitations. *See*, *e.g.*, David Michaels, *Politicizing Peer Review: The Scientific Perspective*, *in* RESCUING SCIENCE FROM POLITICS, *supra* note 197, at 219.

205. *See* David L. Faigman, *Scientific Realism in Constitutional Law*, 73 BROOK. L. REV. 1067, 1083 (2008) (suggesting that the ordinary churn of the scientific method can counter biases that may have entered the process). We don't mean to suggest, however, that somehow science is a pure discipline immune from societal pressures or forces.

206. For useful context on the hacked emails from the Climatic Research Unit (CRU), see *The CRU Hack*, REALCLIMATE, http://www.realclimate.org/index.php/archives/2009/11/the-cru-hack (last visited Nov. 11, 2010).

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.[207] If mud is being thrown, lawyers know how to respond.[208] In addition, law reviews are a fine place to engage in normative debates, which will always be significant even when the descriptive account is agreed upon.

In sum, to those who attack the implicit bias findings as junk science, we disagree on the merits, based on the evidence. If we must cite a single study, it is the predictive validity meta-analysis by Anthony Greenwald and colleagues. That study found that for the forty-seven studies that focused on intergroup behavior in socially sensitive domains, the IAT better predicted behavior than self-reported measures.[209] Moreover, the effect sizes among the studies in this domain were highly homogeneous. And in all nine categories of studies in the meta-analysis, there was significant incremental predictive validity for IAT measures. Whether such evidence is persuasive to lawmakers and policymakers should be determined by the long-run consensus of the scientific community, publishing in scientific journals based on painstakingly collected data, not in punchy editorials[210] or the law reviews. In the meantime, we should be on the lookout for double standards, on all sides, to discern whether something besides just a love of "sound science" is driving the objections.

B. "Hardwired" Resignation

A very different objection agrees wholeheartedly that we are not colorblind—but with the following twist: Implicit biases are hardwired, and

---

207. *Cf.* Faigman, *supra* note 205, at 1083 (arguing that lawyers should "take responsibility for identifying bias where it occurs in empirical research").

208. *See, e.g.*, Bagenstos, *supra* note 178, at 480 (arguing that the critique made by Gregory Mitchell and Philip 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, 1670 (2003) (noting that "a large number of straw men were born and killed in the construction of [Gregory] Mitchell's arguments"); Jeffrey J. Rachlinski, *The Uncertain Psychological Case for Paternalism*, 97 NW. U. L. REV. 1165, 1167 n.18 (2003) (describing Gregory Mitchell's strawman technique as a "cheap academic stunt").

209. *See* Greenwald et al., *supra* note 102, at 28.

210. *See* Karen Lee Torre, *Race Theory Rubbish*, CONN. L. TRIB., June 21, 2010, *available at* http://www.ctlawtribune.com/getarticle.aspx?ID=37500 (complaining of "junk social science"); Amy Wax & Philip E. Tetlock, Op-Ed., *We Are All Racists at Heart*, WALL ST. J., Dec. 1, 2005, at A16. For a response, see Jerry Kang, *The Situation of 'Common Sense'*, SITUATIONIST (July 6, 2010, 12:01 AM), http://thesituationist.wordpress.com.

there's nothing we can do about them. After all, if we have an implicit attitudinal preference for flowers over insects, perhaps these and other biases come from hundreds of thousands of years of natural selection. And, if these associations are immutable, we might as well resign ourselves to our biological, evolutionary, or genetic limitations. This objection will likely be supported with arguments and tropes drawing on sociobiology, evolutionary psychology, and behavioral genetics.

Hardwired resignation objections typically involve two components: strict determinism (one could not have done otherwise) and some genetic or evolutionary origin story for that determinism. The second component—the origin story—grabs all the attention.[211] But the first component—strict determinism—is what's important. Most people (who are not professional philosophers) believe that if we could not have done otherwise, we should not be held morally or legally responsible. By contrast, if one could do otherwise, then we should decide *what* to do and be held accountable for our decisions. This strategy makes sense even if implicit biases were at some point of human history evolutionarily adaptive.[212] To be sure, the origin story may help us reach our desired ends more efficaciously. But that is a "*how* to do it" question more than a "*what* to do" question.

Having focused on the question of strict determinism, we want to clarify the surrounding empirical picture. First, there is great individual variability in implicit biases.[213] Some people show strong bias, but others demonstrate minimal bias or even score 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

211. *See, e.g.*, Jeremy A. Coyne, *Of Vice and Men: The Fairy Tales of Evolutionary Psychology*, NEW REPUBLIC, Apr. 3, 2000, at 27 (reviewing RANDY THORNHILL & CRAIG T. PALMER, A NATURAL HISTORY OF RAPE: BIOLOGICAL BASES OF SEXUAL COERCION (2000) (discussing the evolutionary origin of rape)).

212. 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.

213. The standard deviations shown in Table 1, *supra* Part I.B.1, disclose both prevalence of implicit bias *and* substantial variance. Also, some people are more dispositionally motivated to be nonprejudiced, a tendency that moderates implicit social cognitions. People motivated to be nonprejudiced 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 & SOC. PSYCHOL. 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 & SOC. PSYCHOL. 835 (2002). For a case where motivation was related to a reaction time, but not a physiological measure of bias, see Eric J. Vanman et al., *Racial Discrimination by Low-Prejudiced Whites: Facial Movements as Implicit Measures of Attitudes Related to Behavior*, 15 PSYCHOL. SCI. 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 & SOC. PSYCHOL. BULL. 1332 (2004).

simply to contest crude claims about homogeneous biases predestinating our actions in some gross mechanical way.

Second, even if implicit biases themselves cannot change, the causal link between biases and behavior can be disrupted through procedural and structural reforms. In other words, even if we cannot be fully cognitively colorblind, we can be behaviorally colorblind in various settings. 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.[214] In other hiring contexts, one could imagine "blinding" the names of all resumes in initial sorts.[215] One cannot favor Joshua over Juan even implicitly if neither name appears on the vita.[216]

Third, and most important, implicit biases are highly responsive to environmental exposure and experience.[217] 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 more 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).[218] 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.[219]

Such findings raise the possibility of purposeful training with countertypical datasets. For instance, repeated exposure to a dataset that associates Black faces with positive words has decreased implicit bias, as measured by priming instruments.[220] The implicit stereotype associating BLACKS with *Athleticism* can similarly be dissipated, especially among participants highly motivated not to

---

214. *See* Claudia Goldin & Cecilia Rouse, *Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians*, 90 AM. ECON. REV. 715, 717, 725 (2000).

215. *See* Kang & Banaji, *supra* note 2, at 1092–101 (outlining other strategies that individuals and firms could take to help disrupt the causal chain).

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

217. *See generally* Irene V. Blair, *The Malleability of Automatic Stereotypes and Prejudice*, 6 PERSONALITY & SOC. PSYCHOL. REV. 242 (2002) (literature review).

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

219. *Id.* at 822–23.

220. 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 in implicit bias persisted for two days following exposure to the Black–Positive Words pairing. *See* Michael A. Olson & Russell H. Fazio, *Reducing Automatically Activated Racial Prejudice Through Implicit Evaluative Conditioning*, 32 PERSONALITY & SOC. PSYCHOL. BULL. 421 (2006).

be biased.[221] And practice can attenuate the shooter bias of both police officers and lay samples.[222]

In conclusion, we are unpersuaded by any hardwired resignation objection to adopting legal and social response to implicit bias. The current evidence rejects strict determinism, in cases that range from *Grace's Hire* to *Skip's Encounter*.[223] Accordingly, the law should not ignore the behavioral realist call to account and simply throw up its hands in resignation.

C. "Rational" Justification

If hardwired resignation sounded somewhat wistful, this final objection is aggressive and defiant. It contends that implicit "biases" accurately reflect reality; accordingly, it is rational to act based on them. For example, if we associate ASIAN FACES more than WHITE FACES to the stereotype *Foreign*, it is because that association reflects basic immigration demographics. This may help explain why Asian American Yale students demonstrated this bias as strongly as their White peers.[224] Similarly, if we associate WOMEN more with *Family* than with *Work*, that is because the association reflects basic socioeconomic realities and the division of labor within family units. This helps explain why women seem to hold this bias more strongly than men. According to this objection, "bias" is just a judgmental label for accurate probabilities.

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

221. *See* B. Michelle Peruche & E. Ashby Plant, *Racial Bias in Perceptions of Athleticism: The Role of Motivation in the Elimination of Bias*, 24 SOC. 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).

222. Racial shooter bias in a police simulation decreased after repeated exposure to pairs of stimuli where ethnicity was unrelated to criminality. *See* Plant et al., *supra* note 154, at 149, 153.

223. 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.

224. *See* Devos & Banaji, *supra* note 44.

> their convenience and dignity, but cabdrivers too are entitled to care about their property and safety.[225]

What's new here is the claim that not only are explicit biases accurate but implicit biases are as well. For instance, in an article sharply criticizing the implicit bias literature, Gregory Mitchell and Philip Tetlock write:

> Although reliance on stereotypes can lead to biased beliefs . . . such reliance can also sometimes satisfy technical definitions of individual rationality (that is, stereotypes may have predictive value, and stereotype-driven bias should not be conflated with irrationality or animus.)
>
> [continuing in footnote:] 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 rational if one's race-based stereotypes reflect true differences in base rates of criminality across racial groups.[226]

The "it's rational" justification comprises two claims. The descriptive claim contends that the strength of implicit associations in our brains accurately reflects the state of the world.[227] The normative claim contends that acting on the basis of accurate pictures of the world is rational, and thus morally and legally justified.[228]

---

225. DINESH D'SOUZA, THE END OF RACISM: PRINCIPLES FOR A MULTIRACIAL SOCIETY 252–53 (1995).

226. Mitchell & Tetlock, *supra* note 86, at 1036 & n.41. *See also* Arkes & Tetlock, *supra* note 12, at 258 (suggesting that mental associations measured as implicit social cognitions may in fact be "accurate statistical associations rather than unwarranted conclusions").

227. Mitchell & Tetlock, *supra* note 86, at 1085 ("[I]mplicit prejudice, as now conceived, labels perfectly rational reactions to existing socioeconomic conditions as prejudiced.").

228. *See id.* at 1087 ("[T]he 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.").

More recently, Mitchell and Tetlock have clarified that they have never argued "that rational discrimination should be legal." *See* Gregory Mitchell & Philip E. Tetlock, *Facts Do Matter: A Reply to Bagenstos*, 37 HOFSTRA L. REV. 737, 739 (2009). They write:

> For the record—again—we do not defend the view that rational or unintentional discrimination should be legal or that illegal discrimination must involve animus . . . . But we see no value in basing any model of discrimination—no matter how nobly intentioned—on flawed social science or basic research with no demonstrated external validity for real work settings. That was the overriding message of our earlier article.

Id. at 740.

1. Descriptive Accuracy

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

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 underestimation of some perceived attribute as compared to the "real" measure?), valence accuracy (does this over- or underestimation depend on whether the trait is positive or negative?), and dispersion accuracy (is there an over- or undergeneralization that misestimates the variance of the attribute?).[229]

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 (or probabilities) would have to be mistaken. But there is substantial evidence of dissociation: implicit attitudes and stereotypes differ substantially from explicit attitudes and stereotypes.[230] Which set, then, is accurate? Second, where do these base rates come from? In the domain of race, they come more through vicarious than

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

230. As Table 1 shows, explicit biases are typically far smaller than their implicit counterparts, as measured in standard units. (A standard unit measures magnitude in terms of its number of standard deviations.). These data come from Nosek et al., *supra* note 33.

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 et al., *A Meta-Analysis on the Correlation Between the Implicit Association Test and Explicit Self-Report Measures*, 31 PERSONALITY & SOC. PSYCHOL. BULL. 1369 (2005). That is, explicit and implicit biases are somewhat related, but not highly so.

The best understanding is that implicit and explicit measures tap separate but related sets of cognitions. *See* Greenwald & Banaji, *supra* note 11; Timothy D. Wilson, *A Model of Dual Attitudes*, 107 PSYCHOL. REV. 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 fifty-six out of fifty-seven different attitude objects. *See* Brian A. Nosek, *Moderators of the Relationship Between Implicit and Explicit Evaluation*, 134 J. EXPERIMENTAL PSYCHOL.: GEN. 565 (2005).

through direct experiences with racial others,[231] and there's little reason to think that we are fed accurate, unbiased information through popular culture.[232] Indeed, recent findings have provided further evidence that television can be seen as transmitting something like Trojan horse viruses that exacerbate implicit biases against racial minorities.[233] Third, we tend to see illusory correlations, which arise when two salient events (minority + negative event) are noticed together, which leaves a larger impression on our memories and prompts us to overestimate the frequency of their connection.[234] Fourth, although the memory studies are conflicted, we seem to have preferential recording and recall of stereotype-consistent data.[235]

With respect to valence accuracy, we have substantial motivational desires to exaggerate good traits of ingroups and bad traits of outgroups.[236] 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.[237]

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; they are about liking/disliking and 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 Jews—and who knows the reasons why?—is that somehow accurate? If one

231. For a discussion of direct versus vicarious experiences with racial others, see Jerry Kang, *Cyber-Race*, 113 HARV. 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 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).

232. *See* ROBERT M. ENTMAN & ANDREW ROJECKI, THE BLACK IMAGE IN THE WHITE MIND 49 (2000) (suggesting that vicarious experiences dominate the construction of mainstream culture); Kang, *supra* note 14, at 1563–64.

233. *See* Max Weisbuch et al., *The Subtle Transmission of Race Bias Via Televised Nonverbal Behavior*, 326 SCIENCE 1711 (2009). After analyzing eleven popular TV shows, the researchers found more negative body language (nonverbal behavior) toward Black characters, as compared to White characters of similar status. The researchers controlled for all relevant factors, for example, by removing audio and cropping out the target character from the video to avoid demand effects. *See id.* at 1712. In addition, the researchers found a positive correlation between increased exposure to nonverbal bias (measured by self-reports of television viewing patterns regarding these eleven shows) and higher IAT scores of bias. *See id.* (r=0.28; p=0.047). Finally, they showed that exposure to TV clips that showed pro-White nonverbal behavior produced higher implicit bias scores as well as self-reported bias scores. *See id.* The correlations between exposure and implicit bias scores, in two studies, were r=0.25; p=0.05 and r=0.36; p=0.04.

234. *See* Kang, *supra* note 14, at 1564.

235. *See* Judd & Park, *supra* note 229, at 112 (describing tendencies to overly attend to confirming evidence and underuse disconfirming evidence).

236. *See id.* at 112–13.

237. *See* Kang, *supra* note 14, at 1565.

shows a preferential attitude toward one's ingroup—even if randomly assigned to made-up teams[238]—again, in what ways is that accurate?

b. Rational Processing

Even if we are accurate in our probabilities, we must next process them accurately. It is only a slight exaggeration to suggest that fifty years of hard-nosed scientific inquiry belies any claim of strict rationality.[239] 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 any such computations.[240] Moreover, our motivations to justify the self and the groups we belong to slant how we use or fail to use base-rate information.[241] 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. By contrast, negative attributes of ingroups are viewed as malleable, contingent, and a result of environment or bad luck. Everything is flipped for positive attributes.[242]

Further, empirical evidence suggests that these biases can alter behavior even in ways that are self-detrimental. Recall the resume study that demonstrated discrimination against Lakisha over Emily. That study found that "[a] White name yields as many more callbacks as an additional eight years of experience on a resume."[243] We tend to focus on the harm to the individual, but from the firm's perspective, this evidence shows inefficient, non-profit-maximizing behavior.

---

238. *See supra* note 37.

239. ALEN NEWELL & HERBERT A. SIMON, HUMAN PROBLEM SOLVING (1972); Amos Tversky & Daniel Kahneman, *Judgments Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124 (1974); *see* Banaji et al., *supra* note 64, 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).

240. *See, e.g.*, Aron K. Barbey & Steven A. Sloman, *Base-Rate Respect: From Ecological Rationality to Dual Processes*, 30 BEHAV. & BRAIN SCI. 241, 251–52 (2007) (providing a summary of empirical findings on Bayesian reasoning).

241. *See* Eric Luis Uhlmann, Victoria L. Brescoll & David Pizarro, *The Motivated Use and Neglect of Base Rates*, 30 BEHAV. & BRAIN SCI. 284 (2007).

242. *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, *supra* note 14, at 1566 n.418 (describing this phenomenon).

243. Bertrand & Mullainathan, *supra* note 76, at 992.

Here's further evidence of irrational behavior, using a novel methodology: Eugene Caruso and colleagues[244] had participants choose partners in a trivia game based on traits that were relevant to success (such as IQ and prior experience) and traits irrelevant to success (such as weight, as signaled by a photograph). By using conjoint analysis,[245] the researchers were able to estimate that although "participants reported that weight was the single least important factor in their choice,"[246] 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 IQ points[247] to work with the thinner teammate.[248] In another study with similar methodology, undergraduates were willing to trade away $3249 (22 percent of the salary range of options) to work for a man instead of a woman.[249] Leaving IQ points or money on the table, especially when one explicitly reports that weight and gender don't matter, is hardly rational.[250]

2. Normative Justification

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 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

244. In a conjoint analysis design, participants rate objects (in this case, game partners) that differ on different dimensions. Systematically varying these dimensions with one another across many trials can estimate the relative influence on people's preferences of each factor, such as weight, IQ, or experience with the game. *See* Eugene M. Caruso, Dobromir A. Rahnev & Mahzarin Banaji, *Using Conjoint Analysis to Detect Discrimination: Revealing Covert Preferences From Overt Choices*, 27 SOC. RECOGNITION 128, 131–32 (2009) (describing the method).

245. For an introduction to conjoint analysis, see BRYAN K. ORME, GETTING STARTED WITH CONJOINT ANALYSIS: STRATEGIES FOR PRODUCT DESIGN AND PRICING RESEARCH (2d ed. 2006).

246. *See* Caruso et al., *supra* note 244, at 133.

247. 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). *Id.* at 134.

248. The overweight teammate was much heavier than the thin teammate. Pretesting estimated the overweight teammate to be about 97 lbs heavier (at M=237 lbs) than the thin teammate (at M=140 lbs). *See id.* at 132. If the effect is linear, which may or may not be the case, this translates to 1.27 IQ points for each ten pounds of weight.

249. *See id.* at 136 (citing Dobromir A. Rahnev et al., Conjoint Analysis: A New Method of Investigating Stereotypes (2007) (unpublished manuscript, on file with author)).

250. 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.

(blindspot) errors.[251] Put bluntly, "accurate enough" in *Skip's Encounter*—shoot or don't shoot—is not the same question as "accurate enough" on a friend's lunch order—pepperoni or mushroom. Nothing about rationality tells us how we should evaluate the significance of these Type I and Type II errors.

Even deeper normative anxieties lurk about. Procedurally, if we are making judgments based on probabilities of uncertain accuracy and moral significance, actors should at least publish these probabilities for both the perceiver and targets to appreciate.[252] 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, aged, and educated White man), then shouldn't these data be publicly articulated and vetted? There is no such transparency today.

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 on the radical Left: It is repeated by scholars such as Steven Pinker in his popular rejection of the *Blank Slate*.[253]

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, 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 the ground."[254] Judge Alex Kozinski provides a concrete example:

> Assume you are an anglo homeowner who lives in an all-white neighborhood. Suppose, also, that you harbor no ill feelings toward minorities. Suppose further, however, that some of your neighbors persuade you that having an integrated neighborhood would lower property values and

251. 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.

252. *See* Banaji et al., *supra* note 64, at 285. This recommendation applies most powerfully to state actors, who have greater obligations to be transparent than private actors.

253. *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.").

254. 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); Bagenstos, *supra* note 178, at 486 (pointing out that antidiscrimination laws do not exempt "rational" discriminations). This may be more true of antidiscrimination statutes than of the Constitution. For a discussion of constitutional ambivalence regarding "accurate" or "rational" discrimination, see Jack M. Balkin & Reva B. Siegel, *The American Civil Rights Tradition: Anticlassification or Antisubordination?*, 58 U. MIAMI L. REV. 9, 16–17 (2003).

> that you stand to lose a lot of money on your home. On the basis of that belief, you join a pact not to sell your house to minorities. Have you engaged in intentional racial and ethnic discrimination? Of course you have. Your personal feelings toward minorities don't matter; what matters is that you intentionally took actions calculated to keep them out of your neighborhood.[255]

In sum, 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 for accuracy but also the normative case for why much of current ethical and legal understandings ought to be overturned. That isn't and shouldn't be easy.[256]

CONCLUSION

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 colorblind society. The only exceptions are either hate criminals or intransigent universities that continue to 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 character, not by their social categories. In other words, we see through colorblind lenses. 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.

This convenient story is, however, disputed. In the past, within legal scholarship, it was disputed by passionate retelling of lived experiences.[257] But these narratives were resisted as mere anecdotes, not much different than the purely

---

255. Garza v. Cnty. of Los Angeles, 918 F.2d 763, 778 n.1 (9th Cir. 1990) (Kozinski, J., concurring in part and dissenting in part).

256. 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.

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

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 v. Jackson Board of Education*,[258] 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."[259]

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—for better and worse. The data force us to see through the facile assumptions of colorblindness.[260]

A commitment to behavioral realism could guide the science's incorporation into the law writ large in sensible, nonhysterical ways. But strong objections will be filed, sometimes on the merits and sometimes a bit beyond. What will come to be, we cannot see—only hindsight is 20/20. But by providing clear data and vocabulary, we hope to have clarified the stakes and the future terms of engagement. At the least, we will have accelerated the substantive debate about implicit bias, colorblindness, and the law past caricatures.

258. 476 U.S. 267 (1986).

259. *Id.* at 276 (plurality opinion) (emphasis added).

260. There are numerous anecdotal accounts of how taking the IAT shook the person's self-confidence of colorblindness or egalitarianism. For an account from a federal district court judge, see Bennett, *supra* note 165, at 150 ("I was eager to take the test. I knew I would 'pass' with flying colors. I didn't."). There is also evidence that the mere display of scientific information increases persuasiveness, at least in certain contexts. *See, e.g.*, Deena Skolnick Weisberg et al., *The Seductive Allure of Neuroscience Explanations*, 20 J. COGNITIVE NEUROSCIENCE 470, 475 (2008) (finding that extraneous neuroscience explanations increased how satisfying an explanation was to novice and student readers, but not to experts); David P. McCabe & Alan D. Castel, *Seeing Is Believing: The Effect of Brain Images on Judgments of Scientific Reasoning*, 107 COGNITION 343, 349–50 (2008) (explaining how brain images "appeal to people's intuitive reductionist approach"). We are merely describing what scientists have observed. Of course, we seek relevant, accurate scientific explanations, not superfluous, inaccurate ones.