2007

Implicit Bias, "Science," and Antidiscrimination Law

Samuel R Bagenstos

Implicit Bias, "Science," and Antidiscrimination Law

*Samuel R. Bagenstos**

I. Introduction

In recent years, scholars of antidiscrimination law have increasingly focused on the problem of "implicit" or "unconscious" bias.[1] They have pointed to an expanding mass of evidence from experimental psychology that appears to demonstrate the pervasiveness of unconscious bias based on race, gender, and other legally protected characteristics, evidence that raises troubling questions about the effects of such bias on legally relevant behaviors. For example, research using such tools as the Implicit Association Test (IAT)—"which assesses bias by measuring the speed with which an individual associates a categorical status (such as black or white) with a given characteristic or description (such as good or bad)"[2]—has found that "[w]hite Americans, on average, show strong implicit preference for their own group and relative bias against African Americans."[3] Studies show that whites have similar biases against "other ethnic minority groups such as Latinos, Jews, Asians, and non-Americans," as well as "the elderly, and women."[4] Interestingly, the studies also show that minorities and women often harbor the same implicit biases about their own groups that whites and men harbor against them.[5]

Legal scholars have used that psychological evidence to support a number of doctrinal proposals. For example, Linda Krieger and Susan Fiske have

[1] For reviews of this literature, see Samuel R. Bagenstos, *The Structural Turn and the Limits of Antidiscrimination Law*, 94 Cal. L. Rev. 1, 5–7 & nn.10–20 (2006); Jerry Kang, *Trojan Horses of Race*, 118 Harv. L. Rev. 1489, 1497–1535 (2005). The implicit bias literature obviously builds on the classic works on unconscious bias and antidiscrimination law. *See generally* Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 Stan. L. Rev. 1161 (1995); Charles R. Lawrence III, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 Stan. L. Rev. 317 (1987).

[2] Bagenstos, *supra* note 1, at 6. To take the IAT yourself, visit the IAT website, https://implicit.harvard.edu/implicit/demo/takeatest.html.

[3] Nilanjana Dasgupta, *Implicit Ingroup Favoritism, Outgroup Favoritism, and Their Behavioral Manifestations*, 17 Soc. Just. Res. 143, 147–48 (2004). For a good general overview of the research, see Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 Cal. L. Rev. 945 (2006).

[4] Dasgupta, *supra* note 3, at 147–48 (citations omitted).

[5] *See id.* at 149.

used implicit bias research to criticize the "honest belief rule" and the "same actor inference"—two doctrines employed by many lower courts in addressing claims of intentional discrimination under Title VII of the Civil Rights Act of 1964.[6] In addition, a number of scholars, most notably Jerry Kang and Mahzarin Banaji, and Christine Jolls and Cass Sunstein, have argued that psychologists' implicit bias findings provide a firm justification for affirmative action programs in employment.[7] Other scholars have argued that those findings justify reorienting employment discrimination law to take a more structural approach.[8] Most controversially, Ian Ayres has suggested that IAT scores might "be used as a criterion for hiring both governmental and nongovernmental actors" to reduce the prevalence and effects of implicit bias.[9]

The legal academic literature has not been uniformly supportive of these recommendations. Writing before most of the evidence from the IAT had been reported, Amy Wax argued that "extending the framework created by existing antidiscrimination statutes to cover unconscious workplace disparate treatment is not a good idea because it is unlikely to serve the principal goals of a liability scheme—deterrence, compensation, insurance—in a cost effective manner."[10] In my previous piece on this issue, I agreed that the psychological findings present significant problems for the antidiscrimination project. But I argued that responding to implicit bias requires moving "beyond the generally accepted normative underpinnings of antidiscrimination law."[11] As a result, I (despairingly) thought it unlikely that legislatures would adopt, and courts would carry out, effective responses to the problem.[12]

---

[6] *See* Linda Hamilton Krieger & Susan T. Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 Cal. L. Rev. 997, 1027–52 (2006). Under the "honest belief rule," if an employer honestly believed that there was a nondiscriminatory reason for an adverse employment action, that belief can defeat a finding of discriminatory intent even if there was, in objective fact, no such reason. The "same actor inference" is the inference that if the person who fires an employee is the same one who hired that employee, the firing could not have been based on discriminatory intent.

[7] *See* Christine Jolls & Cass R. Sunstein, *The Law of Implicit Bias*, 94 Cal. L. Rev. 969, 978–88 (2006); Jerry Kang & Mahzarin R. Banaji, *Fair Measures: A Behavioral Realist Revision of "Affirmative Action,"* 94 Cal. L. Rev. 1063 (2006). Michael Selmi made a similar point some years earlier. *See* Michael Selmi, *Testing for Equality: Merit, Efficiency, and the Affirmative Action Debate*, 42 UCLA L. Rev. 1251, 1283 (1995).

[8] *See* Tristin K. Green, *Discrimination in Workplace Dynamics: Toward a Structural Account of Disparate Treatment Theory*, 38 Harv. C.R.-C.L. L. Rev. 91, 95–99 (2003); Susan Sturm, *Second Generation Employment Discrimination: A Structural Approach*, 101 Colum. L. Rev. 458, 460 & n.4 (2001). For a critique of structural proposals, see generally Bagenstos, *supra* note 1.

[9] Ian Ayres, Pervasive Prejudice? Unconventional Evidence of Race and Gender Discrimination 424–25 (2001).

[10] Amy L. Wax, *Discrimination as Accident*, 74 Ind. L.J. 1129, 1132–33 (1999).

[11] Bagenstos, *supra* note 1, at 3.

[12] *Id.* at 34–40. Linda Krieger, in her early work on unconscious bias, sounded similarly skeptical notes. *See* Krieger, *supra* note 1, at 1244–47.

Now, the arguments for using antidiscrimination law to respond to implicit bias face a new, more fundamental challenge. Gregory Mitchell and Philip Tetlock contend, in a recent piece, that the psychological research purporting to demonstrate the pervasiveness of implicit bias "fails to satisfy key scientific tests of validity."[13] Working through numerous studies of implicit bias, Mitchell and Tetlock contend that those studies fail tests of construct validity, internal validity, statistical validity, and external validity.[14] Mitchell and Tetlock ultimately conclude that attaching the label of "science" to implicit bias research is "more honorific than descriptive,"[15] and that the increasing acceptance of that research sets the stage "for an epistemic disaster of minor-epic proportions."[16]

Mitchell and Tetlock make some effective points. In particular, they offer significant reasons for advocates of implicit bias research to be cautious before making sweeping statements that "science" compels one or another proposed legal reform or that only "hypocrisy and self-deception" could justify opposition to it.[17] They also offer good reason to reject the suggestion that employers should use the Implicit Attitude Test to evaluate the implicit attitudes of candidates for managerial or hiring positions.[18]

But, as I hope to show in this Essay, Mitchell and Tetlock's argument does not at all undermine the case for taking account of implicit bias in antidiscrimination policy. Even if one accepts every "scientific" critique they offer of the implicit bias literature—and there is substantial dispute within psychology on some of those critiques[19]—the case for using the law to respond to the problem of implicit bias remains strong. In the end, many of Mitchell and Tetlock's critiques of implicit bias research rest, not on any scientific ground, but on normative assumptions about what kinds of

---

[13] Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 OHIO ST. L.J. 1023, 1023 (2006).

[14] *See id.* at 1056–1115.

[15] *Id.* at 1029.

[16] *Id.* at 1118.

[17] Mitchell and Tetlock single out a passage in which, they claim, Kang and Banaji do just that. *See id.* at 1029. The passage is: "If there is any value judgment embedded in behavioral realism besides those intrinsic to the scientific method, it is a second-order commitment against hypocrisy and self-deception. The law views itself as achieving just, fair, or at least reasonable results. If science reveals that the law is failing to do so because it is predicated on erroneous models of human behavior, then the law must transparently account for the gap instead of ignoring its existence." Kang & Banaji, *supra* note 7, at 1065. I do not believe that Mitchell and Tetlock's is the fairest interpretation of this passage. There, Kang and Banaji are making a general point about how the law ought to respond to science; they are not making any claim about what science shows. I certainly agree, however, that an appeal to "science" is a significant theme of Kang and Banaji's article. In my view, appeals to "science," of the type in which both Kang and Banaji and Mitchell and Tetlock engage, very often mask significant normative disputes. For an elaboration of the point in another antidiscrimination context, see Samuel R. Bagenstos, *The Americans with Disabilities Act as Risk Regulation*, 101 COLUM. L. REV. 1479 (2001).

[18] *See* Mitchell & Tetlock, *supra* note 13, at 1115.

[19] *See infra* note 23.

discrimination the law should seek to prevent and punish.[20] In particular, they rest on a very narrow view, based on notions of individual fault, that the law should prohibit only discrimination that results from self-conscious, irrational animus. That narrow view may resonate politically, but antidiscrimination doctrine and theory have consistently rejected it.[21]

Mitchell and Tetlock's argument is thus best understood, not as a scientific critique of implicit bias research, but as an argument about the normative bases for antidiscrimination law. That argument thus confirms my point that the major obstacle for advocates who seek to retool the law to address implicit bias is the widely accepted set of understandings about the goals of antidiscrimination law.[22] Such advocates must therefore focus their efforts as much on developing the normative case for responding to implicit bias as on developing the scientific case that implicit bias exists.

My argument proceeds as follows. In Part II, I show that Mitchell and Tetlock's arguments do not call into question the case for using antidiscrimination law to respond to implicit bias. In Part III, I offer some thoughts about the normative challenges that remain for advocates of using antidiscrimination law in such a way. In Part IV, I present a brief conclusion.

II. The Resilient Case for Using Antidiscrimination Law To Respond to Implicit Bias

Mitchell and Tetlock amass what seems, cumulatively, to be an overwhelming phalanx of arguments against the science that lies behind proposals for using the law to combat implicit bias. When unpacked, however, nearly all of those arguments rest on a particular set of normative views about the kinds of bias to which antidiscrimination law ought properly to respond. Generally speaking, that view is a narrow one that treats discrimination as a wrong perpetrated by a discriminator who acts self-consciously and irrationally. But advocates of using the law to respond to implicit bias do not take that narrow view. To the contrary, they understand discrimination as a social problem that—whether or not it reflects the "fault" of any individual discriminator—has systematically harmful effects on the life chances of members of particular socially salient groups. Under that broader view of the problem of discrimination, the "scientific" evidence that Mitchell and Tetlock dismiss remains highly relevant and telling. Although they ostensibly attack the "science" of implicit bias research, Mitchell and Tet-

---

[20] In an earlier piece, Richard Banks and colleagues pointed out that criticisms of the race IAT often rest on normative disagreements about the nature of bias. *See* R. Richard Banks et al., *Discrimination and Implicit Bias in a Racially Unequal Society*, 94 Cal. L. Rev. 1169, 1186–87 (2006). My argument goes further and pinpoints some of the ways in which Mitchell and Tetlock's argument rests on such normative disagreements.

[21] For elaboration of the point, see generally Samuel R. Bagenstos, *"Rational Discrimination," Accommodation, and the Politics of (Disability) Civil Rights*, 89 Va. L. Rev. 825 (2003).

[22] *See* Bagenstos, *supra* note 1, at 34–40.

lock's real target is the normative view of antidiscrimination law as reaching beyond acts reflecting the individual fault of the discriminator.

In this Part, I elaborate that point. In Section A, I show that many of Mitchell and Tetlock's criticisms of implicit bias research rest on the question-begging assumption that implicit bias is not a distinct phenomenon—that implicit bias cannot be real if it is not connected to overt and self-conscious prejudice. In Section B, I consider the possible non-"prejudice" explanations Mitchell and Tetlock offer for the studies' findings of implicit bias. Even if one accepts those explanations, I contend that they do not at all undermine the case for treating implicit bias as a serious social problem to which antidiscrimination law should respond. In Section C, by both reviewing the arguments addressed in the first two sections and introducing additional arguments made by Mitchell and Tetlock, I show that their objection to using the law to respond to implicit bias ultimately rests, not on any "scientific" objection, but on an individual-fault-based normative understanding of antidiscrimination law. This does not mean that the case for using the law to respond to implicit bias has been established. Mitchell and Tetlock point out some real questions that need to be answered in the implicit bias literature. And, given the significant consequences to individuals of using the IAT or other measures of implicit bias as a screening device for employment, their arguments should give employers pause before they use those measures in that way.[23] But it does mean that Mitchell and Tetlock have not even come close to delivering a fatal "scientific" blow to the implicit bias antidiscrimination program.

---

[23] For example, the link between implicit bias and discriminatory behavior that actually denies opportunities needs further development. *See* Mitchell & Tetlock, *supra* note 13, at 1067–69 (arguing that the studies show, at most, that implicit bias leads to "micro-level 'discriminatory' behaviors" that are *de minimis*). As workplaces evolve into more flexible, less hierarchical organizations, there is reason to believe that micro-level discriminatory behaviors will increasingly aggregate to cause denials of opportunities, *see* Bagenstos, *supra* note 1, at 11, but the "reason to believe" is not yet hard proof. For discussions of the proof that exists, see Greenwald & Krieger, *supra* note 3, at 953–55; Kang & Banaji, *supra* note 7, at 1072–75. The "'shooter-bias' studies," *see infra* note 34, also provide evidence of a link between implicit bias and deprivation of opportunities. Mitchell & Tetlock, *supra* note 13, at 1068 n.146. At most, however, concern with the link between implicit bias and behavior calls for "future predictive validity studies" that address "the relationship between implicit associations and subjective evaluations of disadvantaged groups." *Id.* at 1069. The relatively low correlation between the two most common methods of measuring implicit bias—the IAT and priming—also raises questions that need to be resolved, *see id.* at 1060–61, although even Mitchell and Tetlock acknowledge that the size of the correlation is the subject of debate within the psychological community, *see id.* at 1061 & n.128 (citing William A. Cunningham et al., *Implicit Attitude Measures: Consistency, Stability, and Convergent Validity*, 12 PSYCHOL. SCI. 163, 167 (2001)). Mitchell and Tetlock also raise important questions about the scaling of measures of reaction speed, *see id.* at 1091–93, and the importance of statistical outliers in the conclusions drawn by implicit bias studies, *see id.* at 1103–05. But these points call for more study of implicit bias measures, not for rejection of the implicit bias program.

A. Begging the Question: Is Implicit Bias a Distinct Phenomenon?

Many of Mitchell and Tetlock's criticisms of implicit bias research rest on the assumption that implicit bias reflects explicit prejudicial attitudes. Thus, they contend that implicit bias research lacks construct validity because the "empirical evidence" regarding the correlation between implicit bias and explicit prejudicial attitudes is "mixed."[24] Similarly, they contend that the implicit bias literature fails to "distinguish between automatically activated associations that, if called to people's attention, they would endorse and those associations that, if called to people's attention, they would categorically reject."[25] People may have negative, race-based implicit associations for a variety of reasons that do not stem from personal prejudice; those associations may instead reflect sympathy or awareness of cultural stereotypes and depressing realities. Such alternative explanations, Mitchell and Tetlock argue, "challenge the characterization of associations as tapping into racial attitudes—at least attitudes in the commonsense view that the attitudes imply an evaluative preference that, when brought to people's attention, they endorse and are even prepared to justify under appropriate conditions."[26]

These arguments do not undermine the case for using law to respond to implicit bias. Scholars who advocate using the law in that way have strongly urged that implicit biases are meaningfully distinct from explicit attitudes.[27] Thus, "even people who express strongly egalitarian attitudes often show significant implicit biases."[28] And, even if those biases are called to their attention, the urge to rationalize them may be as strong as the desire to get rid of them.[29] To say that the concept of implicit bias lacks validity because implicit bias does not correlate empirically with explicit prejudice is therefore to assume the very conclusion that implicit bias scholars seek to challenge—that any "real" bias must be reflected in expressed attitudes (or the attitudes that would be expressed if the matter were called to one's attention).[30]

---

[24] Mitchell & Tetlock, *supra* note 13, at 1062.

[25] *Id.* at 1084.

[26] *Id.* at 1080.

[27] *See, e.g.*, Kang, *supra* note 1, at 1512.

[28] Bagenstos, *supra* note 1, at 7.

[29] *See id.* at 7–9.

[30] *See* Mahzarin R. Banaji et al., *No Place for Nostalgia in Science: A Response to Arkes and Tetlock*, 15 Psychol. Inq. 279, 280 (2004) (responding to an earlier version of the argument made by Mitchell and Tetlock). Relying on an article by Blanton and Jaccard, Mitchell and Tetlock state that one of the advocates of implicit bias research, Mazharin Banaji, has taken two diametrically opposed positions on the question by treating a lack of correlation between implicit and explicit biases as evidence of validity in one article and then treating an observed correlation between the two as evidence of validity in another. *See* Mitchell & Tetlock, *supra* note 13, at 1062 n.130 (citing Hart Blanton & James Jaccard, *Arbitrary Metrics Redux*, 61 Am. Psychol. 62, 66 (2006)); *see also id.* at 1095 (describing implicit prejudice advocates as "caught between a rock and a hard place" on this issue). It is not, I think, consistent with the implicit bias theory to treat a correlation with explicit prejudices

For the same reason, Mitchell and Tetlock's asserted "[d]ramatic dis-confirmation[ ]" of the implicit bias theory—that "one study demonstrated that African-Americans actually preferred to interact with people classified as implicit racists"[31]—is neither dramatic nor a disconfirmation. A key point of implicit bias, to proponents of the theory, is that it is invisible.[32] That minorities—many of whom may themselves share those biases[33]— cannot detect implicit biases in others is not in any way a refutation of the theory. It would be a refutation only if one assumed the very point that implicit bias advocates seek to challenge—that such biases are visible.

B. A Narrow Theory of "Prejudice"

Relatedly, much of Mitchell and Tetlock's argument presumes that only a narrow sort of prejudice ought to be legally relevant. They suggest that, to the extent even blacks are shown to exhibit a bias against blacks, such bias can hardly be the sort of prejudice that ought to concern us. Thus, they dismiss the findings of "shooter-bias" studies, in which participants are more likely to shoot unarmed black targets and refrain from shooting armed white targets,[34] because black participants exhibited the same anti-black bias as did whites.[35] And they contend that "if the theoretical logic of the IAT requires that Jesse Jackson be classified as prejudiced against Blacks, then absurdity cannot be far off."[36]

But that is hardly a scientific refutation of the implicit bias studies. As a logical matter, it is certainly possible that minority group members may be biased against members of their own group; it is hardly scientific to dismiss the possibility out of hand. Indeed, both experience and anti-discrimination doctrine make clear that the possibility is far from merely theoretical. Aside from such cultural clichés as the self-hating Jew,[37] antidis-crimination law and lore are full of stories of intentional discrimination by minority and female supervisors against members of their own race and sex.[38] Such acts of "self-hating discrimination" can have nearly as harm-

---

as validating the theory. But the major thrust of implicit bias research—including the over-whelming bulk of the research published by Banaji—treats the concept as one that is dis-sociated from explicit attitudes. At most, Mitchell and Tetlock's point is a minor "gotcha" that does not undermine the implicit bias theory.

[31] *Id.* at 1065 (citing J. Nicole Shelton et al., *Ironic Effects of Racial Bias During Interracial Interactions*, 16 PSYCHOL. SCI. 397, 401 (2005)).

[32] *See* Bagenstos, *supra* note 1, at 8 (discussing the argument by implicit bias scholars that such bias instantiates through "invisible cognitive processes").

[33] For further discussion of this point, see *infra* text accompanying notes 35–38.

[34] For discussion of these studies, see Kang, *supra* note 1, at 1525–26.

[35] *See* Mitchell & Tetlock, *supra* note 13, at 1068 n.146.

[36] *Id.* at 1085–86 n.204. Studies of implicit bias do show that minorities share the same biases as do whites, but to a substantially lesser degree than do whites. *See* Dasgupta, *supra* note 3, at 149.

[37] *See generally* SANDER L. GILMAN, JEWISH SELF-HATRED: ANTI-SEMITISM AND THE HIDDEN LANGUAGE OF THE JEWS (1986).

[38] For discussions in the law reviews, see Charles A. Sullivan, *Circling Back to the Ob-*

ful an effect on employment opportunities for minorities and women as can "other-hating discrimination" that occurs at the hands of white male supervisors. What should and does matter, for legal and policy purposes, is the fact that an individual is discriminated against; the race or gender identity of the discriminator is immaterial.

Mitchell and Tetlock also contend that "[t]here is strong evidence that psychological processes aside from out-group hostility can artificially inflate and otherwise distort scores on implicit measures such as the IAT."[39] They then proceed to list a number of possible explanations for high implicit bias scores that do not reflect hostility toward minorities: figure-ground asymmetry (i.e., that "greater familiarity with one ethnic-racial group (e.g., Whites over Blacks) drives at least part of the race IAT effect"),[40] stereotype threat (i.e., that the fear of being labeled a bigot will drive some people to behave in a manner that appears bigoted),[41] sympathy (i.e., the prospect that high IAT scores "are rooted more in compassion or guilt about the predicament of African Americans than in hostility or contempt"),[42] and knowledge of the prejudice of others.[43] They offer similar explanations for the findings, in some implicit bias studies, that high IAT scores correlate with more awkward interactions with members of minority groups. Mitchell and Tetlock posit that shame about the way blacks have been treated in this country, or perhaps "social awkwardness stemming from lack of experience with members of other ethnic-racial

vious: The Convergence of Traditional and Reverse Discrimination in Title VII Proof, 46 WM. & MARY L. REV. 1031, 1085 (2004) ("In short, intra-racial or intra-gender discrimination occurs, if more rarely than cross-racial discrimination."); Kenji Yoshino, *Assimilationist Bias in Equal Protection: The Visibility Presumption and the Case of "Don't Ask, Don't Tell,"* 108 YALE L.J. 485, 512 (1998) ("Self-hatred is a common response to [social] stigmatization, as exemplified by the instances of the self-hating black, Jew, or homosexual."). *See* Castaneda v. Partida, 430 U.S. 482, 499 (1977) ("Because of the many facets of human motivation, it would be unwise to presume as a matter of law that human beings of one definable group will not discriminate against other members of their group."); Rinesmith v. Cent. County Fire & Rescue, 156 Fed. Appx. 856, 859 (8th Cir. 2005) ("That two members of the Board of Directors were women does not preclude a finding of sex discrimination. Women may discriminate against women . . . ."); Feingold v. New York, 366 F.3d 138, 155 (2d Cir. 2004) ("We also reject the district court's suggestion that an inference of discrimination cannot be drawn because Feingold was fired by another Jew."); Haywood v. Lucent Tech., Inc., 323 F.3d 524, 530 (7th Cir. 2003) ("Lucent begins by arguing that there should be a presumption of non-discrimination because Foote, who discharged Haywood, is also African-American. It is wrong; no such presumption exists, nor should one be created."); Ross v. Douglas County, 234 F.3d 391, 396 (8th Cir. 2000) ("[W]e have no doubt that, as a matter of law, a black male could discriminate against another black male 'because of such individual's race.'"); *cf.* Kadas v. MCI Systemhouse Corp., 255 F.3d 359, 361 (7th Cir. 2001) (finding it "altogether common and natural for older people . . . to be oblivious to the prejudices they hold, especially perhaps prejudices against the group to which they belong").

[39] Mitchell & Tetlock, *supra* note 13, at 1072.
[40] *Id.* at 1075.
[41] *See id.* at 1079–80.
[42] *Id.* at 1081.
[43] *See id.* at 1084.

groups," could be causing this effect.[44] "Of course," they conclude, "a person experiencing such emotions and displaying awkward nonverbal behaviors is not necessarily prejudiced."[45]

But again, this is not a scientific disagreement so much as a normative one.[46] There is no reason why it should make a difference, from an anti-discrimination law and policy perspective, whether implicit bias reflects "hostility" toward minorities or any of the alternative explanations Mitchell and Tetlock offer. Whatever it reflects, implicit bias can result in behaviors and evaluations that limit the opportunities of minority group members. Racial unfamiliarity ("figure-ground asymmetry" in Mitchell and Tetlock's technical term) exists in the employment setting as well as in the laboratory setting. To the extent that unfamiliarity with members of minority groups in certain work settings leads their white counterparts to engage (even unconsciously) in aversive behavior, that aversion will have the same negative effects on minorities' opportunities as will aversion motivated by hostility.[47] To the extent that stereotype threat is a real risk in the laboratory setting, it can also operate in the employment setting. If whites act in a bigoted way on the IAT because they are afraid of being stereotyped as bigots, as Mitchell and Tetlock's stereotype threat explanation suggests, whites who make employment decisions could just as well act in a bigoted way in those decisions because they are afraid of being stereotyped as bigots.

Sympathy, too, is hardly unproblematic. Paternalistic discrimination has a long history, and it has imposed substantial costs on members of disadvantaged groups. The classic statement of this point is Justice Brennan's opinion in *Frontiero v. Richardson*,[48] which observed that sex discrimination was traditionally "rationalized by an attitude of 'romantic paternalism' which, in practical effect, put women, not on a pedestal, but in a cage."[49] Moreover, work elaborating on the concept of "aversive racism"[50] shows that sympathy, guilt, social awkwardness, and shame have a major effect in reinforcing whites' discrimination against blacks. As Linda Krieger summarizes these findings, "racial ambivalence, normative ambiguity, and

---

[44] *Id.* at 1096.

[45] *Id.* at 1097.

[46] *See* Banaji et al., *supra* note 30, at 282–83 (responding to an earlier version of Mitchell and Tetlock's argument, without emphasizing the normative stakes).

[47] For this reason, I think Mitchell and Tetlock are unfair to Greewald and his colleagues. *See* Mitchell & Tetlock, *supra* note 13, at 1077 n.175.

[48] 411 U.S. 677 (1973).

[49] *Id.* at 684. Racial segregation in schools, too, was justified as being in the interests of blacks. *See* Davis v. County Sch. Bd., 103 F. Supp. 337, 340 (E.D. Va. 1952), *rev'd sub nom.* Brown v. Bd. of Educ., 347 U.S. 483 (1954); Roberts v. City of Boston, 59 Mass. 198, 209–10 (1849).

[50] *See, e.g.*, John F. Dovidio & Samuel L. Gaertner, *On the Nature of Contemporary Prejudice: The Causes, Consequences, and Challenges of Aversive Racism*, *in* CONFRONTING RACISM: THE PROBLEM AND THE RESPONSE 3 (Jennifer L. Eberhardt & Susan T. Fiske eds., 1998).

fear of one's own potential prejudice all serve to amplify white's [sic] dis-crimination against blacks in the giving and requesting of assistance, the evaluation of behavior, physical distancing, and the selection of sanctions for social transgressions."[51] If one cares about responding to conditions that operate as a systemic hindrance to the opportunities of members of minority groups—and that do so through no fault of the minority-group members themselves—one should care about responding to implicit bias, even if that bias is not rooted in hostility.

Finally, Mitchell and Tetlock contend that high implicit bias scores may not reflect prejudice but instead are "perfectly rational reactions to existing socioeconomic conditions."[52] In other words, because "African-Americans have, on average, fewer of the good things in life (high incomes and net worth, college educations, etc.) and more of the bad things in life (higher rates of imprisonment, violence, drug abuse, out of wedlock births, etc.)," it is perfectly rational, and indeed "inevitable," that blackness will come to carry "a substantial network of negative associations."[53]

One might question just how rational these negative associations are. Are they precisely calibrated so that blackness carries negative freight only when and to the extent that the particular black person with whom one deals is statistically likely to be involved in "imprisonment, violence, drug abuse, out of wedlock births, etc."?[54] Even if they are (which Mitchell and Tetlock make no effort to prove), that is entirely irrelevant from the perspective of antidiscrimination law and policy. Antidiscrimination laws do not ex-empt "rational" discrimination from their prohibitions. To the contrary, "[t]he prohibition of rational discrimination is a central component of anti-discrimination doctrine—and it may be the most important aspect of antidiscrimination law on the ground."[55] It is well established, for exam-ple, that Title VII of the Civil Rights Act of 1964 prohibits employers from discriminating on the basis of race or sex even when it is demonstrably more costly to hire workers of a particular race or sex or to assign them to particular positions.[56] Title VII also prohibits employers from discriminating

---

[51] Krieger, *supra* note 1, at 1240 (citations omitted).

[52] Mitchell & Tetlock, *supra* note 13, at 1085.

[53] *Id.* at 1086.

[54] *Id.* Work on cognitive heuristics suggests that they will not be so precisely cali-brated. *See* Cass R. Sunstein, *Probability Neglect: Emotions, Worst Cases, and Law*, 112 Yale L.J. 61, 70–83 (2002).

[55] Bagenstos, *supra* note 21, at 848; *see also* David A. Strauss, *The Myth of Color-blindness*, 1986 Sup. Ct. Rev. 99, 113–16 (1986).

[56] *See* Int'l Union v. Johnson Controls, Inc., 499 U.S. 187, 210 (1991) ("The extra cost of employing members of one sex, however, does not provide an affirmative Title VII de-fense for a discriminatory refusal to hire members of that gender."); Los Angeles Dep't of Water & Power v. Manhart, 435 U.S. 702, 716–17 (1978) (rejecting any "cost justification defense" under Title VII). Mitchell and Tetlock misstate the problem when they say that "the courts permit employers to base decisions on job-relevant attributes correlated with protected-category membership." Mitchell & Tetlock, *supra* note 13, at 1087. Implicit bias, in their rational-association explanation, is more akin to basing decisions on protected-category membership because it is correlated with job-relevant attributes. But that is ex-

in order to satisfy customer preferences—even when customers will, in fact, deny their business to a firm that refuses to discriminate.[57] The law does so because even rational acts of discrimination can aggregate to cause systematic and cumulative disadvantage to members of minority groups.[58] Accordingly, even if implicit biases reflect purely rational discrimination, they remain extremely important for antidiscrimination law and policy. To disagree with that proposition reflects, not a scientific disagreement, but a normative judgment that rational discrimination is not the kind of discrimination the law should try to prohibit.

For the same reasons, Mitchell and Tetlock miss the mark when they assert that implicit bias cannot truly be prejudice because an individual's implicit bias score "depend[s] greatly on the specific stimuli chosen for the sorting task."[59] They ask, rhetorically, "[i]f a relative aversion toward a category only materializes when we use certain types of names or faces—or when we embed those stimuli in certain types of context cues—are we justified in saying people are prejudiced toward the category as a whole or only toward particular exemplars of that category in certain contexts?"[60] The point of antidiscrimination law, however, is not to identify and punish prejudice as an inherent personal quality. The point is to prevent and provide remedies for conduct that deprives minorities of opportunities. Given that goal of antidiscrimination law, the fact that implicit bias may depend on the surrounding stimuli does not suggest that we should *ignore* that bias. Rather, it suggests that law and policy should aim to promote the inclusion in workplaces of the kind of stimuli that reduce implicit bias and its effects. And indeed, that is precisely the conclusion drawn by scholars who advocate using the law to respond to implicit bias.[61] Again, the point is a normative one, not a scientific one.

---

actly what *Johnson Controls* and *Manhart* prohibit.

   [57] *See, e.g.*, Fernandez v. Wynn Oil Co., 653 F.2d 1273, 1276–77 (9th Cir. 1981); Diaz v. Pan Am. World Airways, 442 F.2d 385, 389 (5th Cir. 1971).

   [58] *See* Cass R. Sunstein, *The Anticaste Principle*, 92 Mich. L. Rev. 2410, 2418 (1994) ("[T]he most elementary antidiscrimination principle singles out one kind of economically rational stereotyping and condemns it, on the theory that such stereotyping has the harmful long-term consequence of perpetuating group-based inequalities."). For an extensive effort to justify the prohibition of rational discrimination, see Bagenstos, *supra* note 21, at 848–59.

   [59] Mitchell & Tetlock, *supra* note 13, at 1106. For a similar point, see *id.* at 1113 (noting that photos of black persons "embedded in an egalitarian and positive-affect setting" do not trigger high implicit-bias scores).

   [60] *Id.* at 1107.

   [61] *See, e.g.*, Greenwald & Krieger, *supra* note 3, at 963–64; Christine Jolls, *Antidiscrimination Law's Effects on Implicit Bias*, *in* Behavioral Analyses of Workplace Discrimination (G. Mitu Gulati & Michael Yelnosky eds., forthcoming 2007); Kang & Banaji, *supra* note 7, at 1101–15. A similar point could be made about Mitchell and Tetlock's contention that implicit bias research lacks external validity because workplaces (at least those that embody "best-practices precepts in organizational behavior") differ from experimental settings. *See* Mitchell & Tetlock, *supra* note 13, at 1107–15. External validity is always a difficult problem in experimental psychological research, and implicit bias research is not different. But if "best-practices precepts" can forestall discriminatory ef-

C. An Irrational-Animus Theory of Antidiscrimination Law

Much of the discussion above suggests that what really drives Mitchell and Tetlock's argument is not a scientific disagreement with the implicit bias researchers, so much as a normative disagreement about the appropriate scope and targets of antidiscrimination law. The rational-discrimination point may be especially telling: noting that "the IAT makes it possible to be a Bayesian bigot," they suggest (and cite an article in which Tetlock contends) that the possibility constitutes "a *reductio ad absurdum* of the IAT research program."[62] But it is a *reductio* only if one takes the view that rational discrimination is not a proper target of antidiscrimination law, a point that most civil rights advocates would vigorously deny and that antidiscrimination doctrine categorically rejects.[63] This is not a debate about science; it is a debate about the proper scope of antidiscrimination law.

The assertion that rational discrimination does not count as discrimination suggests that Mitchell and Tetlock view matters through what Alan Freedman called the "perpetrator perspective"—the notion that what matters is whether the person accused of discrimination was at "fault," and not whether the person accused of discrimination actually caused or contributed to group-patterned harm.[64] That perspective is evident throughout the paper. Thus, when considering the possibility that implicit biases might be activated by sympathy, rather than antipathy, toward minority

---

fects of implicit bias, that point suggests that more workplaces should adopt those precepts. It is worth noting that many workplaces do not incorporate those precepts. When workplaces do, it is often a response to legal antidiscrimination mandates. *See, e.g.*, *id.* at 1109 (noting that "hiring and staffing managers at many organizations," unlike laboratory subjects, "have been alerted to . . . the dangers of discrimination" and "are well aware of the need to have adequate and legal justifications for their judgments and decisions"); *id.* at 1110 (noting that managers, unlike laboratory subjects, "often receive clear and repeated admonishment" that they may not "allow job-performance-irrelevant characteristics, such as membership in ethnic-racial groups, to affect their personnel judgments"); *id.* at 1114 (noting "intergroup contact is associated with reduced prejudice, including 'prejudice' at the implicit level") (footnote omitted). To the extent that legal rules encourage adoption of these best-practices precepts—and therefore promote intergroup contact at the workplace—that is an argument *for*, not against, legal intervention in those workplaces that have not adopted them. Relatedly, Mitchell and Tetlock's argument that the "stranger-stranger interactions employed in" implicit bias experiments "do not reflect the types of interactions that give rise to most employment discrimination lawsuits," *id.* at 1068, misses the mark because much of the employment discrimination *problem* involves hiring discrimination. That few hiring suits are filed says more about the barriers to success in hiring-stage litigation than it does about the scope of the discrimination problem. *See* John J. Donohue III & Peter Siegelman, *The Changing Nature of Employment Discrimination Litigation*, 43 Stan. L. Rev. 983, 1015–17 (1991).

[62] Mitchell & Tetlock, *supra* note 13, at 1085 & n.204.

[63] For an excellent argument for the wrongfulness of what the author explicitly calls "Bayesian" discrimination, see Jody D. Armour, *Race Ipsa Loquitur: Of Reasonable Racists, Intelligent Bayesians, and Involuntary Negrophobes*, 46 Stan. L. Rev. 781 (1994).

[64] *See* Alan David Freeman, *Legitimizing Racial Discrimination Through Antidiscrimination Law: A Critical Review of Supreme Court Doctrine*, 62 Minn. L. Rev. 1049, 1052–54 (1978).

groups, Mitchell and Tetlock ask "what moral stance should be taken toward these individuals?"[65] When discussing measures of "modern racism," they scornfully ask "whether people should be labeled racists for believing that many or most obstacles to racial equality are now internal to the Black community."[66] And they show an intense concern for those accused of implicit bias by noting the "dangers of making false accusations of prejudice"[67] and by worrying that the IAT labels individuals with explicit egalitarian views as "implicit bigots on par with children reared in prejudiced households and taught to hold mean-spirited beliefs about minorities and to act out these prejudices."[68]

As Mitchell and Tetlock themselves acknowledge, these concerns are "ultimately a matter of political values, not scientific fact."[69] If one sees antidiscrimination law as limited to punishing individuals who engage in irrational, animus-based discriminatory conduct, then Mitchell and Tetlock's concerns about false accusations of prejudice will resonate. But advocates of using the law to counter implicit bias deny that antidiscrimination law should be about punishment or that it should focus only on irrational animus. Rather, they contend that antidiscrimination law should respond to the harmful effects of the "mechanisms of bias as produced by the current, *ordinary* workings of human brains—the mental states they create, the schemas they hold, and the behaviors they produce."[70] They thus treat implicit bias not as a matter of individual fault, but as a social problem with harmful consequences to which the government must respond. The disagreement is, again, a normative one. It is not a scientific one.

In their conclusion, Mitchell and Tetlock pull back the curtain to reveal the normative agenda their "scientific" arguments really support. They argue that the claims of implicit bias advocates profoundly "shake the ontological foundations of American political culture" because those claims "imply that we cannot achieve equality of opportunity unless we have already achieved equality of result."[71] And they argue that those claims

---

[65] Mitchell & Tetlock, *supra* note 13, at 1083.

[66] *Id.* at 1082 n.190; *see also id.* at 1064 (arguing that explicit measures of "modern, symbolic, and aversive racism often include items that could easily serve as measures of ideological conservatism, traditional values, and the Protestant work ethic"). Note that this criticism reaches beyond implicit bias research to indict much research into *explicit* and *overt* prejudice.

[67] *Id.* at 1118; *see also id.* at 1101 (worrying about "false-accusation rates").

[68] *Id.* at 1088; *see also id.* at 1104 (expressing concern that the IAT research program "is being used to indict the vast majority of the American population as implicitly prejudiced and disposed to discriminate whenever a suitable pretext emerges").

[69] *Id.* at 1101.

[70] Kang & Banaji, *supra* note 7, at 1075 (emphasis added). Indeed, Kang and Banaji specifically disclaim the argument "that implicit bias-induced discrimination should produce the same legal liability as explicit animus-driven discrimination under current equal protection doctrine or federal antidiscrimination statutes." *Id.* at 1077.

[71] Mitchell & Tetlock, *supra* note 13, at 1120.

"stimulat[e] excessive suspicion of Whites among Blacks, suspicion that can crystallize into conspiracy theories that poison race relations."[72]

These are, of course, normative arguments. And there is substantial reason to doubt them. Terms like "equality of opportunity" and "equality of result" have strong political resonance, but they have no clear analytic meaning.[73] To the extent there is a meaningful distinction between the two concepts,[74] the claims of implicit bias scholars seem clearly to fall on the "equality of opportunity" side of the line. Those scholars do not claim that every race has to be equally represented in all professions; they claim that implicit and unconscious biases deny members of minority groups the opportunities to which they would be entitled if the biases were not present. To say that those scholars seek more than equal opportunity requires defining the term so narrowly as to guarantee nothing more than freedom from overt prejudice. If one disagrees that equal opportunity is so limited a concept, the bulk of Mitchell and Tetlock's "scientific" arguments are basically irrelevant to an evaluation of the project of using the law to combat implicit bias.

III. The Normative Challenges That Remain

In the previous Part, I showed that Mitchell and Tetlock's challenges are not so much scientific as normative. If one takes the view that antidiscrimination law should aim solely at punishing individuals who act on irrational racist and sexist animus, their arguments will resonate strongly. But if one takes the view that discrimination is a social problem that entrenches the subordinated and disadvantaged status of particular, socially salient groups, then their arguments will be far less relevant. In some respects, those arguments call for caution and further study before taking the more extreme steps urged by some scholars who write about implicit bias—steps like using the IAT as an employment screening device[75]—but they do not undermine the strong reasons to believe that implicit bias contributes to a serious social problem that denies opportunities to blacks and other minorities in America. Indeed, by emphasizing the context-dependence of implicit bias, their arguments actually lend support to some of the proposals offered by legal scholars who urge a response to implicit bias.[76]

I might just stop there. Mitchell and Tetlock have attempted to deliver a fatal blow to the implicit bias law-reform program, but their blow

---

[72] *Id.* at 1119.

[73] *See generally* David A. Strauss, *The Illusory Distinction Between Equality of Opportunity and Equality of Result*, 34 Wm. & Mary L. Rev. 171 (1992).

[74] *See id.* at 172 ("The distinction between opportunity and result is an unhelpful and misleading way to categorize social institutions.").

[75] *See supra* note 23 and accompanying text.

[76] *See supra* text accompanying note 61.

has largely missed the mark. Let us proceed with the program—with caution, to be sure, but let us proceed.

But things are not so simple. Mitchell and Tetlock's argument likely *will* resonate with many lawyers, judges, and policymakers. The justification that best fits antidiscrimination doctrine is one that focuses, not on the individual fault of the discriminator, but on the social harms of discrimination.[77] But most people—even most policy actors—have not internalized that justification. Legal interventions that go past the individual-fault justification and press against the social-harm one therefore encounter political and judicial backlash.[78] The adverse reaction to the Americans with Disabilities Act, whose requirement of "reasonable accommodation" centrally targets *rational* employer conduct, demonstrates this point.[79]

What I have done in this Essay is therefore only the first step. Science does not defeat the implicit bias law-reform program, but science does not establish the case for that program, either. That program depends on a normative judgment that discrimination is not about fault but about a social problem—a normative judgment that is deeply contested among judges and policymakers. Unfortunately, arguments about implicit bias too often place heavy rhetorical emphasis on appeals to "science," perhaps at the expense of developing the normative arguments for applying antidiscrimination law to that important phenomenon.[80]

To be sure, the results of psychological studies can help to address the normative debate. The "shooter bias" studies, for example, are powerful because of the severe nature of the harm they suggest that implicit bias can cause.[81] It hardly stretches widely accepted normative understandings to suggest that one's race should not be the basis for being shot—even if it is automatic processes of the brain, and not anything that is the "fault" of the officer, that causes a police officer to shoot an unarmed black man.[82] When implicit bias affects actions that are not so immediately and

---

[77] *See generally* Bagenstos, *supra* note 21.

[78] *See* Bagenstos, *supra* note 1, at 40–45.

[79] *See* Linda Hamilton Krieger, *Socio-Legal Backlash*, 21 BERKELEY J. EMP. & LAB. L. 476 (2000).

[80] Kang and Banaji repeatedly seek to invoke the mantle of "science" in a way that, at least rhetorically, seems intended to do normative work, *see supra* note 17, although they also recognize that there are possible normative differences between discrimination actuated by implicit bias and other sorts of discrimination. *See supra* note 70. Linda Krieger recognizes that "even the best insights from the empirical social sciences cannot supply the normative principles needed for substantive lawmaking or resolve the conflicts between competing norms and interests so often implicated in the legislative and judicial processes." Krieger & Fiske, *supra* note 6, at 1007. But her recent work on the issue rhetorically relies heavily on invocations of empiricism and science. *See generally* Greenwald & Krieger, *supra* note 3, *passim*; Krieger & Fiske, *supra* note 6, *passim*. Jolls and Sunstein touch on some of the normative concerns here, but only briefly. *See* Jolls & Sunstein, *supra* note 7, at 992–95.

[81] *See supra* note 34.

[82] *See* Banks et al., *supra* note 20, at 1173–75 (discussing how the shooter bias studies

ultimately consequential, but are instead "the everyday actions of virtu-
ally all of us,"[83] the normative case for responding to that bias becomes
more complex.

So far, the most significant effort to develop a normative response to
problems of implicit bias has come from those scholars who advocate a
structural approach to antidiscrimination law.[84] Invoking the "democratic
experimentalist" theory that merges early twentieth-century American prag-
matism with late twentieth-century Japanese management principles, Susan
Sturm has argued that the normative principles for responding to struc-
tural discrimination problems will best develop through a process in which
those problems are addressed at the local level and information is shared
through professional and other networks.[85] I gave qualified (though de-
cidedly pessimistic) endorsement to that proposal in my earlier piece on
these issues.[86]

But to a large extent, the democratic experimentalist move is a punt.
It remains incumbent on employment discrimination scholars to articulate
and elaborate the normative theory under which the law should respond
to implicit bias. Such theorizing can, of course, inform the efforts of lo-
calized activists who operate in a democratic-experimentalist world. But
the theorizing is important, more fundamentally, to convince policymak-
ers and judges to endorse the legal regimes that can address implicit bias.

My call, then, is for a renewed attention to antidiscrimination *theory*.
As the social and behavioral sciences have gained prominence in the le-
gal academy, much antidiscrimination work has come to explore the eco-
nomics and psychology of discrimination. Work on implicit bias is just a
part of that trend. There is much that this kind of work can do to help illu-
minate the important issues of antidiscrimination law. Just as did eco-
nomic analysis, psychological research into implicit bias can help us bet-
ter understand both the need for and the limits of antidiscrimination law.
If continued implicit bias research shows increasingly strong correlations
between measures of that bias and behaviors that (cumulatively) have signi-
ficant discriminatory effects, that information will be centrally important
to a full understanding of the discrimination problem. But at the same time
scholars continue to develop that psychological research, we must con-
tinue to develop our normative understandings of antidiscrimination law
to fit the findings of that research as well.[87] And that, ultimately, is the

---

were inspired by the widely publicized shooting of Amadou Diallo by New York police
officers).

[83] Bagenstos, *supra* note 1, at 42–43.

[84] *See* Sturm, *supra* note 8, at 473–74, 559–64.

[85] *See id.* at 559–64.

[86] *See* Bagenstos, *supra* note 1, at 45–47.

[87] That, in part, was what I was trying to do in Bagenstos, *supra* note 21. Tristin Green
has also taken up this challenge. *See* Tristin K. Green, *A Structural Approach as Antidis-
crimination Mandate: Locating Employer Wrong*, 60 Vand. L. Rev. (forthcoming 2007),
available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=903791.

lesson of Mitchell and Tetlock's ostensibly scientific challenge to implicit bias research.

IV. Conclusion

Implicit bias research has already informed antidiscrimination scholarship, and it is likely to inform antidiscrimination scholarship even more in the future. But, as Mitchell and Tetlock's argument highlights, the implicit bias law-reform project is incomplete. This is true not so much for scientific reasons (though more work clearly needs to be done) but for normative ones. Mitchell and Tetlock's arguments underscore that point. Those arguments fail as a scientific challenge to implicit bias research because they largely rest on a set of unexpressed normative disagreements with implicit bias scholars. The implicit bias law-reform program still stands, but scholars advocating for that program must do a better job of articulating and defending the normative propositions on which their arguments depend. "Science" will not save antidiscrimination law. To save antidiscrimination law requires articulating and defending the normative principles that justify and guide the application of that law to newly understood forms of bias, whatever they are, and however they are discovered.