

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

# Fair Measures: A Behavioral Realist Revision of "Affirmative Action"

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

| | |
|---|---|
| **Citation** | Jerry Kang and Mahzarin R. Banaji, Fair Measures: A Behavioral Realist Revision of Affirmative Action, 94 Cal. L. Rev. 1063 (2006). |
| **Published Version** | doi:10.2307/20439059 |
| **Accessed** | February 19, 2015 3:50:30 PM EST |
| **Citable Link** | http://nrs.harvard.edu/urn-3:HUL.InstRepos:12220342 |
| **Terms of Use** | This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA |

*(Article begins on next page)*

# Fair Measures: A Behavioral Realist Revision of "Affirmative Action"

Jerry Kang†
Mahzarin R. Banaji††

> Bias both conscious and unconscious, reflecting traditional and unexamined habits of thought, keeps up barriers that must come down if equal opportunity and nondiscrimination are ever genuinely to become the country's law and practice.
> — Justice Ginsburg,
> dissenting in *Adarand Constructors, Inc. v. Pena*[1]

> One thing I have learned in a long life: that all our science, measured against reality, is primitive and childlike—and yet it is the most precious thing we have.
> — Albert Einstein[2]

## INTRODUCTION: A NEW BEGINNING

The term "affirmative action" includes a broad range of policies and practices designed to promote equality in ways not strictly required by

Copyright © Jerry Kang and Mahzarin R. Banaji

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

†† Richard Clarke Cabot Professor of Social Ethics, Department of Psychology, Carol K. Pforzheimer Professor at the Radcliffe Institute for Advanced Study, Harvard University. (Mahzarin_Banaji@harvard.edu) (http://www.peoplefas.harvard.edu/~banaji) (www.implicit.harvard.edu) Helpful research assistance was provided by Tami Kameda, Kathryn Padbury, Jennifer Roche, Gwen Sedney, and the Hugh & Hazel Darling Law Library at UCLA School of Law. We benefited from talks given at Maryland Law School, Georgetown Law Center, UCLA School of Law's Critical Race Studies Program, Yale Law School, and Virginia Law School. For helpful comments, we thank: Ian Ayres, members of the Banaji Lab, Richard Banks, Gary Blasi, Devon Carbado, Dolly Chugh, Jennifer Eberhardt, Richard Fallon, Susan Fiske, Mark Greenberg, Anthony Greenwald, Lani Guinier, Christine Jolls, John Jost, Louis Kaplow, Ken Karst, Sung Hui Kim, Bill Klein, Russell Korobkin, Linda Hamilton Krieger, Tom Newkirk, Kristina Olson, Jeff Rachlinski, Judith Resnik, Russell Robinson, Lee Ross, Michael Selmi, Margaret Shih, Reva Siegel, Gerry Spann, Cass Sunstein, Michael Rip Verkerke, Eugene Volokh, Robin West, Joan Williams, and Noah Zatz. This work was supported in part by Georgetown University Law Center; FAS, Harvard University; Radcliffe Institute for Advanced Study at Harvard University; Russell Sage Foundation; Third Millennium Foundation; UCLA School of Law; and the UCLA Asian American Studies Center.

1. 515 U.S. 200, 274 (1995) (Ginsburg, J., dissenting).
2. ALBERT EINSTEIN, THE EXPANDED QUOTABLE EINSTEIN 261 (Alice Calaprice ed., 2000).

antidiscrimination law alone.[3] Since its inception in the 1960s,[4] affirmative action has produced volumes of moral, legal, and policy arguments both to justify and undermine its very existence. The original framing and subsequent discourse have been premised largely on historical and moral-philosophical arguments, which are now well rehearsed and not especially persuasive to those who disagree. Indeed, we seem to be at a deadlock of policy and principle, resistant to any fundamental reexamination. What might lead us out of this stasis?

We believe that new facts recently discovered in the mind and behavioral sciences can potentially transform both lay and expert conceptions of affirmative action. Specifically, the science of implicit social cognition (ISC) can help us revise the very meaning of certain affirmative action prescriptions by updating our understanding of human nature and its social development.

The science of ISC examines those mental processes that operate without conscious awareness or conscious control but nevertheless influence fundamental evaluations of individuals and groups. As described by Anthony Greenwald and Linda Krieger in this Symposium, evidence from hundreds of thousands of individuals across the globe shows that (1) the magnitude of implicit bias toward members of outgroups or disadvantaged groups is large,[5] (2) implicit bias often conflicts with conscious attitudes, endorsed beliefs, and intentional behavior,[6] (3) implicit bias influences evaluations of and behavior toward those who are the subject of the bias,[7] and (4) self, situational, or broader cultural interventions can correct systematic and consensually shared implicit bias.[8] As disturbing as this evidence is, there is too much of it to be ignored. Moreover, recent discoveries regarding malleability of bias provide the basis to imagine both individual and institutional change.

Behavioral realism takes ISC science seriously. The methodology of behavioral realism forces the law to confront an increasingly accurate

---

3. We use the term "affirmative action" as it is used colloquially. It includes a broad range of policies and practices that are designed to respond to past discrimination, prevent current discrimination, and promote certain societal goals such as social stability or improved pedagogy. Affirmative action programs may be facially race- or gender-neutral (for example, broadcasting widely a particular employment opportunity) or race- or gender-contingent (for example, providing some resource to a woman or racial/ethnic minority under circumstances in which that person would not have received the resource but for that person's status as a woman or minority).

4. *See generally* SAMUEL LEITER & WILLIAM M. LEITER, AFFIRMATIVE ACTION IN ANTIDISCRIMINATION LAW AND POLICY: AN OVERVIEW AND SYNTHESIS (2002); JOHN DAVID SKRENTNY, THE IRONIES OF AFFIRMATIVE ACTION: POLITICS, CULTURE, AND JUSTICE IN AMERICA (1996).

5. Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CALIF. L. REV. 954-58 (2006).

6. *Id.* at 953.

7. *Id.* at 953-54, 961-62.

8. *Id.* at 962-65.

description of human decision making and behavior, as provided by the social, biological, and physical sciences. Behavioral realism identifies naïve theories of human behavior latent in the law and legal institutions. It then juxtaposes these theories against the best scientific knowledge available to expose gaps between assumptions embedded in law and reality described by science. When behavioral realism identifies a substantial gap, the law should be changed to comport with science.[9] If legal actors and policy makers decline to revise the law, they should act transparently and provide the prudential, economic, political, or religious reasons for retaining a less accurate and outdated view.

Behavioral realism is more a methodology than a set of first-order normative commitments or policy preferences. Of course, the methodology relies on assumptions and values inherent in the conduct of modern science, which supposes that the causal processes of the real world exist and operate independent of what we know or think about them, and that the scientific method provides one of the best ways of understanding those causal processes.[10] We further recognize that empirical findings cannot replace values, and by themselves, do not dictate any single course of action.[11] 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.

Using behavioral realism as our legal approach and ISC as the science, we seek to revise the affirmative action conversation. First, we provide a new temporal framing for much affirmative action discourse based on the evidence of pervasive implicit bias. No longer do we have to choose between a backward-looking frame of corrective justice (e.g., compensation for slavery) and a forward-looking frame of utilitarian engineering (e.g., potential pedagogical benefit). Instead, we can now view core "affirmative action" programs as responses to *discrimination in the here and now*. We do not dismiss the significance of a historical view and its moral pull, or the potential benefits in social stability and economic growth

9. We are not suggesting that any such gap necessarily exists everywhere in the law. *See, e.g.*, Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 U. CHI. L. REV. 571, 608-18 (1998) (suggesting that the law accounts for hindsight bias).

10. *Cf.* IAN SHAPIRO, FLIGHT FROM REALITY 8-9 (2005) (identifying these qualities as core commitments).

11. We are also mindful of how political agendas can be cloaked as "mere empirical refinements." Deborah Jones Merritt, *Constitutional Fact and Theory: A Response to Chief Judge Posner*, 97 MICH. L. REV. 1287, 1290 (1999) (criticizing Judge Richard Posner for "disguis[ing] theoretical difference as commitment to empirical fact").

arising from diversity. But we contend that a presentist framing that exposes and responds to pervasive implicit bias—even in those who genuinely believe themselves to be bias-free—provides an independent and compelling case for action.

Second, and closely connected, we update the scientific case for the *mismeasurement of merit*. Critics of affirmative action argue that affirmative action circumvents merit. However, the presence of implicit bias can produce discrimination by causing the very basis of evaluation, merit, to be mismeasured. This insight reframes certain affirmative action programs not as "preferential treatment" but as an opportunity for more accurate measures.

Third, we analyze ISC findings to suggest new approaches to ameliorating the problem of bias, or *debiasing*.[12] Affirmative action has sometimes been credited for producing the sort of integration that reduces stereotypes and prejudice. The mechanism for this benefit is the well-known "social contact hypothesis" (SCH), which social psychologists have refined, complicated, and challenged over the past five decades.[13] ISC suggests experimenting with debiasing mechanisms different from the traditionally recommended peer-to-peer social contact; potential techniques include self-propelled attitude makeovers, mental "contact" through imagery, and exposure to debiasing agents.

In the Conclusion, we suggest a fourth and final insight—a *new ending* for affirmative action. In *Grutter v. Bollinger*, Justice O'Connor suggested a twenty-five year fuse on affirmative action.[14] In our view, however, the lifespan for certain affirmative action programs should be guided by evidence of bias rather than any arbitrary or hopeful deadline. Now that we can measure threats to fair treatment—threats that lie in every mind—such data should be a crucial guide to ending affirmative action. We suggest a terminus when measures of implicit bias for a region or nation are at zero or some rough behavioral equivalent. At this point, implicit bias would align with an explicit creed of equal treatment. It would fulfill collective aspirations to behave in accordance with explicitly held values.

A nomenclature clarification: although we use the term "affirmative action," we find it too freighted to be analytically useful. As we make specific recommendations based on our analysis of ISC, we employ where

12. *See* Christine Jolls & Cass R. Sunstein, Debiasing Through Law (Nov. 18, 2003) (unpublished manuscript, Yale Legal Theory Workshop, on file with authors); *see also* Baruch Fischhoff, *Debiasing*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS & BIASES 423 (Daniel Kahneman et al. eds., 1982) (discussing debiasing to counter certain heuristics and biases); Jerry Kang, *Trojan Horses of Race*, 118 HARV. L. REV. 1491, 1537 (2005) (discussing behavioral realist research agenda in terms of "debiasing" solutions).

13. Thomas F. Pettigrew & Linda R. Tropp, *A Meta-Analytic Test of Intergroup Contact Theory*, J. PERSONALITY & SOC. PSYCHOL. (forthcoming 2006).

14. 539 U.S. 306, 343 (2003) ("We expect that 25 years from now, the use of racial preferences will no longer be necessary to further the interest approved today.").

possible a different term, "fair measures." "Fair" connotes the moral intuition that being fair involves an absence of unwarranted discrimination, by which we mean unjustified social category-contingent behavior.[15] The term also connotes accuracy in assessment. "Measure" has a double meaning as well: measurement and an intervention intentionally taken to solve a problem.

This renaming is substantive not cosmetic. Some fair measures we advocate—for example, anonymous evaluations—would not be construed as affirmative action as much as antidiscrimination. In this sense, fair measures include items not conventionally contained in the affirmative action label. Conversely, some forms of what is now called affirmative action—for example, reparations on a purely corrective justice theory, or racial minority hiring to generate greater firm revenues—could not be strictly justified as a fair measure. Hence, fair measures are both broader and narrower than affirmative action. Our case in favor of fair measures says little for or against other "affirmative action" or social justice interventions outside its purview.

## I
## DISCRIMINATION NOW, SOLUTIONS NOW

### *A. Conventional Thinking: Backward and Forward*

The relationship between the problem of discrimination and the solution of affirmative action is not straightforward. To some, affirmative action principally corrects historical discrimination against subordinated social groups. This *backward-looking* defense of affirmative action, sounding in corrective justice,[16] runs into well-known political and legal

15. To elaborate further, "discrimination" involves different treatment by a perceiver (discriminator) of a target (victim) based on the social category to which the target has been mapped. A combination of stereotypes (cognitive component) about and attitudes (affective component) toward the target's social category causes the perceiver to treat the target differently. In our definition of discrimination, the actor's self-awareness of both the different treatment and its actual causes is irrelevant. Finally, discrimination can range from warranted to unwarranted, as a function of the applicable moral and legal frameworks. For example, in the United States today, we believe that there is wide consensus that discrimination justified on the grounds of White genetic supremacy is *un*warranted both morally and legally. Other cases are in greater dispute, such as discrimination to pursue corrective justice. In this Article, when we use the term "discrimination," we generally mean *unwarranted* discrimination. If we mean otherwise, we signal accordingly.

16. Roughly speaking, corrective justice is the idea that those who have wronged a person have a moral obligation to make amends to the person and make that person whole. The classical citation is to Aristotle, who called this "justice in rectification." *See* ARISTOTLE, NICOMACHEAN ETHICS 125-28 (Terence Irwin trans., Hackett Publishing Co. 1985). Backward looking justifications may also sound in terms of retributive justice, which emphasizes the punishment of the wrongdoer over making the victim whole.

obstacles.[17] For example, our society tends to view discrimination as a species of individual tort. Accordingly, any claim for remedy must—more or less—identify the wrong, the specific perpetrator, and the specific victim.[18] Only after all three have been confidently specified will the law require the specific wrongdoer to provide proportional redress to the specific victim.

Affirmative action as traditionally understood does not fit this narrow model.[19] As for perpetrators, slave-owners are long dead, and those who have inherited advantages are not held directly accountable for what their ancestors did decades or centuries ago.[20] Not surprisingly, Whiteness is not viewed as a corporation that carries its specific debts forward.[21] Further, the beneficiaries of affirmative action today (e.g., recent minority immigrants) are not regarded as the specific victims of the prior discrimination, or even their heirs. Thus, any benefit they receive is decried as "unjust enrichment".[22]

Constitutional law generally reflects these sentiments. Race-conscious affirmative action programs must undergo the same strict scrutiny reserved for Jim Crow laws.[23] Further, the state cannot have the declared objective

---

17. *See, e.g.*, Kathleen M. Sullivan, *Sins of Discrimination: Last Term's Affirmative Action Cases*, 100 HARV. L. REV. 78, 92 (1986) (noting how this framing "invites claims that neither nonvictims should benefit, nor nonsinners pay").

18. *See, e.g.*, Alan David Freeman, *Legitimizing Racial Discrimination Through Antidiscrimination Law: A Critical Review of Supreme Court Doctrine*, 62 MINN. L. REV. 1049, 1052-54 (1978); Alan Freeman, *Antidiscrimination Law: The View From 1989*, 64 TUL. L. REV. 1407, 1412-13 (1990) (describing "perpetrator perspective").

19. *See generally* Kenneth Karst, *The Revival of Forward-Looking Affirmative Action*, 104 COLUM. L. REV. 60, 61-62 (2004) (noting the awkwardness of compensation for past discrimination justification).

20. This argument is made even more specifically about slavery reparations. *See, e.g.*, David Horowitz, *Ten Reasons Why Reparations for Blacks Is a Bad Idea for Blacks—and Racist Too!*, Mar. 12, 2001, http://www.adversity.net/reparations/anti_reparations_ad.htm. For responses, see, e.g., Charles J. Ogletree Jr., *Repairing the Past: New Efforts in the Reparations Debate in America*, 38 HARV. C.R.-C.L. L. REV. 279, 308-12 (2003).

21. According to a CNN/USA Today/Gallup poll conducted in 2002, only 6% of Whites agreed that the United States government should make cash payments to those descendents of slaves. When asked about corporations who profited from slavery, 11% of Whites responded that they should make cash payments to descendents of slaves. When the form of reparations was not cash payment but a scholarship fund, 35% of Whites agreed that corporations who profited from slavery should pay. PollingReport.com, Race and Ethnicity, http://www.pollingreport.com/race (last visited Jan. 16, 2006). For reasons why American society, if not Whiteness per se, must carry its debts forward for the legacy of slavery and discrimination against African-Americans, see Kim Forde-Mazrui, *Taking Conservatives Seriously: A Moral Justification for Affirmative Action and Reparations*, 92 CALIF. L. REV. 683, 715-26 (2004).

22. Of course, White privilege can be seen as the more significant problem of "unjust enrichment." *See generally* Cheryl I. Harris, *Whiteness as Property*, 106 HARV. L. REV. 1709 (1993).

23. *See, e.g.*, Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 227 (1995) (requiring strict scrutiny for race-conscious affirmative action, even if conducted by the federal government); City of Richmond v. J.A. Croson Co., 488 U.S. 469 (1989) (requiring same strict scrutiny for state and local governments).

of remedying general "societal discrimination."[24] Instead, it must only remedy ongoing acts of discrimination or "lingering effects" of prior discrimination[25] evidenced by particularized and reliable legislative, judicial, or administrative findings.[26] In addition, the affirmative action program must be narrowly tailored, which typically requires consideration of race neutral alternatives, careful demarcation of the beneficiary class, limited period of operation, and minimization of burden on those excluded from the affirmative action program.[27] The law prohibits inflexible quotas that maintain a strict racial balance.[28]

In response to such political and legal constraints on backward-looking justifications for affirmative action, liberal proponents have adopted instead a *forward-looking* justificatory frame.[29] These proponents rally around "diversity," praising its pedagogical[30] and quality-of-service benefits.[31] By sidestepping the blame-game and enlarging the class of indirect beneficiaries (e.g., everyone in the classroom benefits from class

---

24. *See, e.g.*, Grutter v. Bollinger, 539 U.S. 306, 323-24 (2003); Wygant v. Jackson Bd. of Educ. 476 U.S. 267, 274 (1986) (Powell, J.) (plurality opinion) ("This Court never has held that societal discrimination alone is sufficient to justify a racial classification."); Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265, 310 (1978) (rejecting interest in remedying societal discrimination for fear of harming innocent third parties); *see also* Girardeau A. Spann, *Constitutionalizing And Defining Racial Equality: The Dark Side of* Grutter, 21 CONST. COMMENTARY 221, 230 n.53 (2004) (tracing doctrinal history of this position, starting from Justice Powell's plurality opinion in *Bakke* to Justice O'Connor's majority opinion in *Grutter*).

25. *See Adarand*, 515 U.S. at 237.

26. *See Croson*, 488 U.S. at 497-508. It is difficult to specify precisely what type of evidence will be deemed adequate. *See generally* JOHN E. NOWAK & RONALD D. ROTUNDA, CONSTITUTIONAL LAW § 14.10, at 804-07 (7th ed. 2004).

27. *See, e.g.*, *Grutter*, 539 U.S. at 341 (explaining that narrow tailoring requires programs that do "not unduly harm members of any racial group"). Significant burdens, such as losing seniority protection in layoffs, have been held to be unconstitutional. *See, e.g.*, *Wygant*, 476 U.S. at 267. *See also* Robert C. Post, *The Supreme Court, 2002 Term—Foreword: Fashioning the Legal Constitution: Culture, Courts, and Law*, 117 HARV. L. REV. 4, 66-67 (2003) (summarizing the narrow tailoring requirements from *Grutter* as requiring: no undue harm of any racial group; serious, good faith consideration of race-neutral alternatives; limitation in time; and individualized consideration).

28. *See, e.g.*, *Bakke*, 438 U.S. at 306-07 (opinion of Powell, J.) (rejecting the goal of "reducing the historic deficit of traditionally disfavored minorities in medical schools and in the medical profession" as impermissible racial balancing); Freeman v. Pitts, 503 U.S. 467 (1992) ("Racial balance is not to be achieved for its own sake.") (school desegregation context).

29. *See, e.g.*, Sullivan, *supra* note 17, at 96-97 (arguing that forward looking frames are "less vulnerable to 'white innocence' challenges and claims of 'nonvictim windfalls'").

30. *See, e.g.*, COMPELLING INTEREST: EXAMINING THE EVIDENCE ON RACIAL DYNAMICS IN HIGHER EDUCATION (Mitchell J. Chang et al. eds., 2003); Mitchell J. Chang et al., *The Educational Benefits of Sustaining Cross-Racial Interaction Among Undergraduates*, 77 J. HIGHER EDUC. 430, 449 (2006) (finding that cross racial interaction increased "openness to diversity," "cognitive development," and "self-confidence," as measured by self-reports).

31. *See* Paul Frymer & John D. Skrentny, *The Rise of Instrumental Affirmative Action: Law and the New Significance of Race in America*, 36 CONN. L. REV. 677, 677 (2004) ("[A]ffirmative action is increasingly being justified not as a remedy to historical discrimination and inequality, but as an instrumentally rational strategy used to achieve the positive effects of racial and gender diversity in modern society.").

diversity), this framing has produced some political traction.[32] In doctrinal terms, this forward-looking frame was precisely the door left open by Justice Powell's plurality opinion in *Bakke*, which emphasized that the pedagogical benefits of a racially diverse classroom are compelling.[33] In *Grutter*, with surprising decisiveness,[34] the Supreme Court preserved and arguably expanded this forward-looking frame[35] by upholding the constitutionality of moderate forms of race-based affirmative action in law school admissions.[36]

However, even this limited adoption of the forward-looking frame has sparked controversy. First, many are skeptical about the true pedagogical value added by diversity in the classroom.[37] Does it really deserve to be called a "compelling" interest? Second, some argue that the diversity justification should operate across the intellectual and political spectra. This would entail valuing more socially conservative, religious, and right-wing representation in the academy.[38] Conservatives proffer the fact that liberals have not agitated for such diversity as evidence that their commitment to "diversity" is insincere.[39] Third, critics claim that forward-looking frames have no limitation principle because one can always conjure up potential policy benefits of a race-conscious distribution of resources. Fourth,

---

32. *Cf.* Eugene Volokh, *Diversity, Race as Proxy, and Religion as Proxy*, 43 UCLA L. REV. 2059, 2060 (1996) (pointing out that diversity "ascribes no guilt, calls for no arguments about compensation").

33. *Bakke*, 438 U.S. at 311-12.

34. Certain lower courts had held that educational diversity was not a compelling interest. *See, e.g.*, Hopwood v. Texas, 78 F.3d 932 (5th Cir. 1996), *overruled in part by Grutter*, 539 U.S. at 322. Also, the Supreme Court had suggested that remedying past discrimination may be the only permitted justification for race-conscious remedies. *See Croson*, 488 U.S. at 493 (plurality opinion).

35. For example, in *Grutter*, the majority gave great weight to the amicus briefs of former military leaders and General Motors, which claimed that diversity produced a more effective combat and work force. *See, e.g.*, Consolidated Brief of Lt. Gen. Julius W. Becton, Jr. et al., as Amici Curiae at 7-9, Grutter (No. 02-241), Gratz v. Bollinger, 539 U.S. 244 (2003) (No. 02-516); Brief of General Motors Corporation as Amicus Curiae at 23-24, Grutter (No. 02-241), Gratz (No. 02-516).

36. *See Grutter*, 539 U.S. at 327-30, 334 (upholding the University of Michigan Law School's affirmative action program as narrowly tailored to further the compelling interest of educational diversity). *But see Gratz*, 539 U.S. at 270-75 (invalidating the affirmative action program used in the University of Michigan's undergraduate admissions).

37. For example, racial diversity probably cannot improve the way that students learn multivariable calculus. Likewise, homogeneous firms may operate more efficiently, at least in the short term. *See* Devon W. Carbado & Mitu Gulati, *The Law and Economics of Critical Race Theory: Crossroads, Directions, and a New Critical Race Theory*, 112 YALE L.J. 1757, 1789-1802 (2003) (book review) (discussing the literature demonstrating such efficiencies).

38. *See, e.g.*, David Horowitz, *In Defense of Intellectual Diversity*, CHRON. OF HIGHER EDUC., Feb. 13, 2004, at B12, *available at* http://chronicle.com/free/v50/i23/23b01201.htm (suggesting that his Academic Bill of Rights would promote intellectual diversity).

39. *See, e.g.*, Gabriel J. Chin, *Bakke to the Wall: The Crisis of Bakkean Diversity*, 4 WM. & MARY BILL RTS. J. 881, 930 (1996) (suggesting that for those who support affirmative action on backward-looking grounds, "the diversity fig leaf exists as a pretext"); James Lindgren, *Conceptualizing Diversity in Empirical Terms*, 23 YALE L. & POL'Y REV. 5 (2005) (reporting comments by Harvard Law Professor Randall Kennedy that "No one really believes in diversity.").

general anxiety pervades the strategy of transforming race or gender into a qualification.[40] In the end, many Americans seem unpersuaded by arguments about past wrongs[41] or promises about future value (often disconnected from the problem of discrimination).[42] This raises the question: how does behavioral realism alter the frame?

### *B. Behavioral Realism: Here and Now*

Most fundamental is the pervasive, replicable, and sometimes large effects of implicit bias in the here and now.[43] Implicit biases are not merely an academic concern, although their discovery has shaped new theories of mental processes.[44] Implicit bias has consequences in the daily activities of our lives. Indeed, on socially sensitive matters such as discrimination, implicit bias scores have greater predictive validity than explicit self-reports.[45] The assumption is that individuals are not necessarily withholding their "true" attitudes and beliefs but rather that they are unable to know the contents of their mind.

To parse the policy implications of this science we must examine the magnitude of bias (how big is it?), its pervasiveness (how many people does it affect?), and its ability to predict real-world behavior (is the bias merely some strength of association in the mind that remains there?).

---

40. *See, e.g.*, Frymer & Skrentny, *supra* note 31, at 722 ("[Instrumental affirmative action] has allowed some forms of diversity to prosper, but in the process, it has weakened the legitimacy of affirmative action to remedy historic discrimination against those most in need. Most importantly, instrumental affirmative action may limit opportunities for minorities in ways that remedial affirmative action does not."). As Frymer and Skrentny point out, the District Court in Patrolmen's Benevolent Ass'n v. City of New York, 74 F. Supp. 2d 321 (S.D.N.Y. 1999) upheld the use of race-specific hiring of police officers to prevent racial unrest, after the brutal abuse of Abner Louima by New York police officers. In this case, however, Black police officers complained that they were consistently assigned to lower status jobs and exposed to greater danger. *See id.* at 335-36.

41. This characterization is descriptive, not normative. The present consequences of past wrongs are enormous. And, we believe it reasonable for American society to act on its moral obligation to respond to these consequences aggressively. *See generally* Forde-Mazrui, *supra* note 21, at 733 (arguing that current disparities between Black and White populations were "proximately caused" by the racism of the past).

42. Forward looking justifications are not always so disconnected. For example, much of the rhetoric in *Grutter* strayed beyond the pedagogical benefit of diversity and emphasized diversity's value in creating a well-integrated society that functioned with less discrimination and also appeared as doing so. *See infra* text accompanying notes 244-245.

43. *See* Greenwald & Krieger, *supra* note 5 at 953-58, 961-62. Implicit bias is a scientific term of art. It refers to the displacement of response along some judgment dimension caused by implicit attitudes or implicit stereotypes. *See* id at . Although implicit bias can be measured in many different ways, a principal technique is to measure the differences in speed of response between alternative pairings of social categories on the one hand and attitudinal valences or stereotypical traits on the other. Implicit and explicit bias cause changes in behavior, which we call discrimination. As already explained, that discrimination may be warranted or unwarranted, legal or illegal. *See supra* note 15 (defining discrimination).

44. *See, e.g.*, Mahzarin R. Banaji, *Implicit Attitudes Can* Be *Measured, in* THE NATURE OF REMEMBERING 117 (Henry L. Roediger et. al. eds., 2001).

45. *See* Greenwald & Krieger, *supra* note 5 at 954 (referencing Poehlman).

Recently, a public website that administers the implicit association test (IAT)[46] has accumulated a large database of well over three million tests, which now provide an answer to the first two questions.[47] For instance, by a conservative estimate, around ninety percent of Americans (and others in the western world), mentally associate negative concepts with the social group "elderly"; only about ten percent show the opposite effect associating elderly with positive concepts. Seventy-five percent of Whites (and fifty percent of Blacks) show anti-Black bias, and seventy-five percent of men and women do not associate female with career as easily as they associate female to family. These results contrast sharply with the views expressed on explicit surveys.[48] These data, as well as the findings in dozens of experiments that meet the criteria of replicability and peer-review, demonstrate that we are not color or gender blind, and perhaps that we cannot be.

Of course, these implicit associations in our minds may lack any behavioral manifestations. However, the recent predictive validity meta-analysis by Andrew Poehlman, Eric Uhlmann, Anthony Greenwald, and

---

46. For a description of the IAT, see *id.* at 952-53. The IAT has been and continues to be studied very carefully. "Importantly, it has been shown that IAT measures are internally consistent, not confounded by participants' overall speed, right or left handedness, or familiarity with IAT stimuli, and are relatively insensitive to methodological factors like the number of target stimuli and trials and the interval between the target stimuli and required response." T. Andrew Poehlman et. al., Understanding and Using the Implicit Association Test: III. Meta-analysis of Predictive Validity 5 (unpublished manuscript, on file with authors). (internal citations omitted) To get a rough sense of the rise in influence of the IAT, we searched the PsychInfo database for "implicit association test." In 1998, we found only three records; in 1999, seven records; in 2000, eighteen records; in 2001, seventy-one records; in 2002, eighty records; in 2003, 140 records; in 2004, 187 records; and as of Dec. 14, 2005, in 2005, 135 records.

47. The data are reported in Greenwald's contribution to this Symposium. Greenwald & Krieger, *supra* note 5 at 957-58. *See also* Brian A. Nosek et. al., *Harvesting Implicit Group Attitudes and Beliefs from a Demonstration Web Site*, 6 GROUP DYNAMICS 101, 105 (2002) (reporting findings from a dataset with N = 192,364). The dataset was created through volunteers completing a test on the Internet, which is not a random sample. However, this sample was far more demographically diverse than the laboratory samples traditionally drawn from college psychology students. Furthermore, the results can be compared against more traditional laboratory data. *See id.*; *see also* Robert Kraut et. al., *Psychological Research Online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet*, 59 AM. PSYCHOLOGIST. 105, 106 (2004) (arguing that an advantage of Internet research is the ability to produce "a large, diverse sample at low cost" and citing the collection of "over 2.5 million responses in tests of implicit attitudes and beliefs" as an example; Nosek et al., *supra*, at 104 (addressing other caveats).

48. From their large Internet dataset, Brian Nosek and colleagues found "implicit biases were notably stronger than their explicit counterparts and were sometimes in contradiction to them." Nosek et. al., *supra* note 47, at 111. For example, explicit measures showed White respondents had a preference for Whites over Blacks, and Black respondents had a strong preference for Blacks over Whites. But on implicit measures White respondents demonstrated a strong preference for Whites over Blacks, and Black respondents had a weak preference for Whites over Blacks. *Id.* at 105; *see also* Greenwald & Krieger, *supra* note 5 at.

Mahzarin R. Banaji[49] indicates that implicit bias correlates with real-world behavior.[50] In this study, the researchers analyzed a total of 224 IAT-behavior correlations, generated from sixty-nine statistically independent samples, drawn from twenty-one peer-reviewed published studies and thirty-one unpublished studies.[51] They found that implicit biases correlated with real-world behaviors like being friendly toward a target, allocating resources to minority organizations, and evaluating job candidates (weighted mean correlation $r = 0.25, p = 10^{-71}$).[52] In other words, those who show a larger bias on the IAT also discriminate more in their behavior.[53]

Jeff Rachlinski, et al., cautions that many of the behavioral measures that were correlated were intermediary steps to some final decision.[54] In other words, even if high implicit bias correlates with stiff body language, that does not *necessarily* demonstrate disparate treatment in the final selection. Still, influencing intermediary steps likely produces different end-results, at least in close cases.[55] Also, we point out that correlations have been found between implicit bias and ultimate decisions, such as hiring recommendations[56] and funding decisions.[57] In research produced since the meta-analysis, additional correlations between implicit bias and ultimate decisions have been found.

For example, Jonathan Ziegert and Paul Hanges had participants act as managers instructed to evaluate job candidates based on paper

---

49. *See* Poehlman, *supra* note 46. For an explanation of what a meta-analysis is and its substantial benefits, see R. Rosenthal & M.R. DiMatteo, *Meta-Analysis: Recent Developments in Quantitative Methods for Literature Reviews*, 52 ANN. REV. PSYCHOL. 59 (2001).

50. The researchers defined "behavioral measure" as "any measure of a physical action, judgment, decision or physiological reaction." Poehlman, *supra* note 46, at 5.

51. This was the entire universe of relevant studies that the researchers could locate through research in the PsycInfo database, Google, and email contact with a social psychology mailing list asking for unpublished and in press studies as of June 20, 2003. *See id.* By considering both published and unpublished studies, the researchers could check whether there was some publication bias that favored large effect sizes. To the contrary, the r values were higher in the unpublished studies (r = .29) as compared to the published ones (r = .21).

52. Table 1 provides a list of the behaviors with which IAT measures have been found to correlate.

53. For an application of ISC to managerial decision making, see Mahzarin R. Banaji et. al., *How (Un)ethical Are You?*, HARV. BUS. REV., Dec. 2003, at 3; Max H. Bazerman et. al., *When Good People (Seem to) Negotiate in Bad Faith*, NEGOTIATION (Harv. Bus. Sch. Publ'g, Boston, Mass.), Oct. 2005, at 1.

54. *See* Jeffrey J. Rachlinski, et al., Does Unconscious Bias Affect Trial Judges? (manuscript on file with author) (cited with permission).

55. For a more detailed explanation of the positive feedback loops in awkward body language that lead to worse interviews "on the merits," see Kang, *supra* note 12, at 1524-25.

56. *See, e.g.*, Laurie A. Rudman & Peter Glick, *Prescriptive Gender Stereotypes and Backlash Toward Agentic Women*, 57 J. SOC. ISSUES 743, 757 (2001). In this experiment, in one condition, a gender IAT correlated with a hireability index computed on the basis of survey responses to three questions: "that (1) they would interview the applicant for the job, (2) they would personally hire the applicant for the job, and (3) the applicant would be hired for the job." *Id.* at 751-52.

57. *See* Laurie Rudman et al., Minority Member's Implicit Stereotypes and Attitudes (unpublished manuscript, on file with authors).

dossiers.[58] The dossiers were designed to be comparable in quality, with race (Black or White) randomly assigned. In a hiring condition in which the president of the firm signaled his preference for a White hire,[59] implicit bias correlated significantly with disparate evaluations. When no such preference was expressed, there was no correlation. This suggests that the institutional environment influences whether implicit biases are behaviorally manifested.[60]

Skeptics may question the external validity of laboratory-based studies[61] where the respondents are typically college psychology students who may neither take the experiments seriously nor consider fully the consequences of their actions. But consider a recent study conducted by Alexander Green, Dana Carney, and Mahzarin R. Banaji, which examined how medical interns made diagnoses as a function of race.[62] Two hundred and ninety-one medical interns in the Boston and Atlanta metropolitan areas were randomly assigned to view, read symptom profiles, and make diagnosis and treatment recommendations for a hypothetical Black or White patient. Consistent with the prevalence of coronary artery disease (CAD) in Black and White Americans, Black patients were more likely to be diagnosed with CAD than White patients.

However, treatment with state of the art Thrombolytic Therapy was given *equally* to both Black and White patients thereby creating a greater discrepancy between diagnosis and treatment for Black than White patients. The most highly biased medical interns as measured by the IAT were also more likely to treat White patients with Thrombolytic Therapy, despite their own diagnoses of Black Americans' higher likelihood of CAD. The greater disparity between diagnosis and treatment for Blacks relative to Whites was best accounted for by a path model showing that IAT bias led to a stereotype that Blacks were stubborn and noncompliant and therefore likely to refuse treatment. In sum, even when the participants

---

58. *See* Jonathan C. Ziegert & Paul J. Hanges, *Employment Discrimination: The Role of Implicit Attitudes, Motivation, and a Climate for Racial Bias*, 90 J. APPLIED PSYCHOL. 553, 556 (2005).

59. The president's memo asked the manager to consider education and experience. But in the "racial bias" condition, the memo included the following paragraph: "Given that the vast majority of our workforce is White, it is essential we put a White person in the VP position. I do not want to jeopardize the fine relationship we have with our people in the units. Betty (the outgoing vice president) worked long and hard to get those folks to trust us, and I do not want her replacement to have to overcome any personal barriers." *Id.* at 558. This manipulation seems unrealistic because such preferences are no longer written down; that said, the outlandishness of the request should have worked *against* finding any behavioral correlation.

60. *See id.* at 559, 561.

61. For a defense of laboratory experimentation as the tool for secure generalizations, see Mahzarin R. Banaji & Robert G. Crowder, *The Bankruptcy of Everyday Memory*, 45 AM. Psychologist. 1185 (1989).

62. Alexander Green, Dana Carney & Mahzarin R. Banaji, Measuring Physicians' Implicit Biases: A New Approach To Studying Root Causes Of Racial/Ethnic Disparities In Health Care (unpublished manuscript, on file with authors).

(doctors) were making recommendations in a serious context and were arguably subject to strong demand effects to demonstrate that they were colorblind, they still engaged in disparate treatment that correlated with their implicit biases.

This empirical demonstration of implicit bias and its consequences enables a new temporal framing for affirmative action not based on the past or future, but the *present*. This presentist framing—to borrow Kathleen Sullivan's words—does not deny that the past is in need of "redemption."[63] Nor does it deny the benefits of diversity, which experiments in social cognition can help identify and measure. It does, however, foreground the evidence of widespread implicit bias here and now.

A presentist framing avoids the temporal problems of a backward-looking frame. The passage of time disrupts chains of causality and weakens both moral and legal claims for correction.[64] But the presentist approach does not look to the past. It also does not highlight "institutional racism," which skeptics refute as unfalsifiable and as merely some regrettable but not unjust disparate impact inevitable in market competition. Instead, it points to 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. Obviously, both history and societal factors play a crucial role in providing the content of those schemas, which are programmed through culture, media, and the material context.[65] But the presentist approach does not rely on some amorphous racism brooding "out there"; it focuses instead on the bias measurable within individuals.[66]

A presentist framing also avoids problems with forward-looking "diversity" justifications of affirmative action. These justifications were politically attractive—arguably necessary—because we, as a society, lost political consensus on the magnitude of bias and discrimination that persisted. With evidence from ISC, the forward-looking frame becomes optional. We do not need to argue about the empirical benefits of diversity—

---

63. Sullivan, *supra* note 17, at 98.

64. Richard Epstein calls this a "wasting asset" with a "built-in time fuse." Richard A. Epstein, *A Rational Basis for Affirmative Action: A Shaky But Classical Liberal Defense*, 100 U. MICH. L. REV. 2036, 2039 (2002). For arguments why past inequalities continue to manifest themselves today, see, e.g., GLENN C. LOURY, ANATOMY OF RACIAL INEQUALITY 23-30 (2002) (discussing self-reinforcing stereotypes); Michelle Adams, *Intergroup Rivalry, Anti-Competitive Conduct and Affirmative Action*, 82 B.U. L. REV. 1089, 1117-22 (2002) (applying lock-in theory to explain the inequalities between Blacks and Whites in education, housing, and employment markets); Daria Roithmayr, *Barriers to Entry: A Market Lock-In Model of Discrimination*, 86 VA. L. REV. 727, 743-48 (2000) (providing overview of lock-in theory, drawing on antitrust law and concepts).

65. There is little reason to think that racial schemas are significantly hardwired. *See, e.g.*, Kang, *supra* note 12, at 1531-35 (responding to "correction is impossible" objection); Andreas Olsson et al., *The Role of Social Groups in the Persistence of Learned Fear*, 309 SCIENCE 785, 787 (2005) (rejecting simplistic evolutionary biology story).

66. *See* Shankar Vedantam, *See No Bias*, WASH. POST, Jan. 23, 2005, at W12 (discussing "thumbprint of the culture on our minds") (quoting Mahzarin R. Banaji).

although we can. We do not need to explain why such real-world benefits trump the supposed moral or constitutional imperative of colorblindness—although we can. Instead, by demonstrating discrimination *now*, this fire can be fought with narrowly tailored fire. Put another way, color consciousness in the form of pervasive implicit bias is what requires color consciousness in the form of prevention and remedies.

This reframing has political implications. It speaks to the many Americans who are willing to adopt fair measures that take race and gender explicitly into account only to stop and prevent unwarranted discrimination on the basis of those very attributes. This reframing also has doctrinal consequences, on both the compelling interest and narrow tailoring prongs of an equal protection analysis. For example, *Grutter* held that educational diversity was constitutionally compelling. We do not know, however, whether a new Supreme Court will trim this finding or expand it beyond the domain of higher learning.[67] Regardless, it is indisputable that *responding to discrimination* is a compelling interest not limited solely to the field of education.[68]

We make two clarifications. First, responding to discrimination should be a constitutionally compelling interest regardless of whether explicit or implicit bias actuates the discrimination.[69] Those who argue otherwise must confront the science that demonstrates the existence and real-world consequences of implicit bias. Given this evidence, they bear the burden to show why these harms, whether they be couched in terms of inefficiency or unfairness, should be categorically disregarded simply because their causes operate beneath our self-awareness.[70] Ignorance is not always a defense.

---

67. *Grutter* invites a broader reading that goes beyond just the field of education and into the forward-looking benefits of a diversified elite in the military as well as the business worlds. *See, e.g.*, Karst, *supra* note 19, at 60-61, 67. Justice Scalia publicly fretted over just this fate. *See* Grutter, 539 U.S. at 348 (Scalia, J., dissenting) (suggesting that diversity might be used in employment context). Post-*Grutter*, the Seventh Circuit Court of Appeals, per Judge Richard Posner, held that a diversified police force was a compelling interest. *See* Petit v. City of Chicago, 352 F.3d 1111, 1114-15 (7th Cir. 2003).

68. We focus on constitutional arguments in the body of the text. But there are statutory implications too, for example in the application of Title VII of the 1964 Civil Rights Act. As described in greater detail *infra* text accompanying note 255, Title VII may not tolerate voluntary race-conscious strategies justified on forward-looking "diversity" goals. In sharp contrast, a presentist goal of stopping discrimination is compatible with the purpose of Title VII. Again, the reframing makes a tangible difference.

69. In the Title VII context, we think that responding to discrimination is consistent with the statute regardless of whether explicit or implicit bias drives the behavior. There is some case law support for this position. *See, e.g.*, Watson v. Fort Worth Bank and Trust, 487 U.S. 977, 990-91 (1988) (noting unconscious bias being a problem even if intentional discrimination is not occurring); McDonnell Douglas Corp. v. Green, 411 U.S. 792, 801 (1973) ("Title VII tolerates no racial discrimination, subtle or otherwise.").

70. It would be as if bruises from an easy-to-see punch should be legally cognizable, but cancer from hard-to-see benzene exposure must be categorically ignored. We understand that in a tort case, the former is easier to prove than the latter. But that hardly means that society should not have laws and policies that forbid benzene dumping or decrease its unnecessary production.

We are not arguing that discrimination caused by implicit bias should be *equally* problematic as that caused by consciously endorsed explicit bias. Even though the latter is more offensive to equality, responding to the former remains a compelling interest. Second, given our presentist reframing, *responding* to discrimination means not only *remedying* present acts of discrimination but also *preventing* discrimination that is likely to occur without some proactive action.[71] Indeed, this is one of the original meanings of "affirmative action."[72] In sum, preventing (not only remedying) discrimination caused by implicit (not only explicit) bias should be considered a compelling interest.

To this conclusion, we incorporate the arguments made by Ian Ayres and Frederick Vars, that public actors can adopt affirmative action in a specific market in order to remedy *private* discrimination within that same market.[73] Drawing on *Croson*,[74] they argue persuasively[75] that the state is not artificially constrained to combat the bias solely of its own employees and agents. Instead, the state can adopt narrowly tailored measures that provide better treatment in the public sector to counter the worse treatment in the private market.[76] With this addition, we reach the following doctrinal conclusion: The state's preventing discrimination by itself or remedying discrimination by certain delimited private actors[77] is a constitutionally compelling interest regardless of whether the discrimination is caused by explicit or implicit bias.

We are *not* arguing that implicit bias-induced discrimination should produce the same legal liability as explicit animus-driven discrimination under current equal protection doctrine or federal antidiscrimination statutes. That question rests beyond our project. Instead, our legal analysis

---

71. Michael Yelnosky has persuasively argued that the preventative justification for voluntary affirmative action is in accord with Title VII. *See* Michael J. Yelnosky, *The Prevention Justification for Affirmative Action*, 64 OHIO ST. L.J. 1385 (2003).

72. *See* William W. Van Alstyne, *Affirmative Actions*, 46 WAYNE L. REV. 1517, 1527-29 (2000) (explaining that Executive Order 11246, issued under the Presidential administrations of Kennedy and Johnson, was directed at federal contractors and consisted of "precautionary and preventive measures" because of the "concern that were they not taken, some racial discrimination might otherwise occur").

73. *See* Ian Ayres & Fredrick E. Vars, *When Does Private Discrimination Justify Public Affirmative Action?*, 98 COLUM. L. REV. 1577, 1581 (1998). Ayres and Vars provide the example of a city engaging in affirmative action for minority-owned subcontractors because of clear evidence that private contractors discriminate against minority subcontractors. *See id. See also* Kenneth L. Karst, *Private Discrimination and Public Responsibility: Patterson in Context*, 1989 SUP. CT. REV. 1, 44 (noting Justice O'Connor in *Croson* "ma[de] clear that [a] city can accept its share of the public responsibility for remedying private discrimination, [by] using its spending powers to remedy discrimination in the local construction industry).

74. *See* City of Richmond v. J.A. Croson Co., 488 U.S. 469, 491-92 (1989) (a state "has the authority to eradicate the effects of private discrimination within its own legislative jurisdiction").

75. We do not undertake any separate defense of their argument. Also, this addition is fully severable from the arguments we have made up to now.

76. *See* Ayres & Vars, *supra* note 73, at 1611-19.

77. *See id.* at 1615-33.

considers only whether courts should deem a *voluntary* adoption of a fair measure that counters implicit bias-induced discrimination to be a "compelling interest" when opponents legally challenge the measure.

Now that we have discussed the compelling interest prong, we shift to the narrow tailoring discussion. Courts have rejected the backward-looking goal of remedying general societal discrimination partly because of the difficulty of narrowly tailoring any response to such an immense and pervasive problem.[78] Similar concerns about narrow tailoring have derailed forward-looking objectives that courts have criticized as open-ended social engineering without adequate tethers to restrain its operations.[79] By contrast, as we demonstrate throughout the paper, fair measures that target discrimination now can be more objectively designed, implemented, and delimited in scope and duration.

### C. *A Better Model of Discrimination*

If behavioral realism reorients us to consider discrimination here and now, one might reasonably ask why standard antidiscrimination law does not suffice? In other words, if the compelling interest is to prevent discrimination, why do we need fair measures beyond antidiscrimination laws?

Linda Hamilton Krieger ably addressed these questions,[80] so we add only a few points. Lawmakers developed traditional antidiscrimination law in ignorance of ISC generally and implicit bias specifically. The basic components of traditional antidiscrimination law are (1) ex ante commands not to discriminate, and (2) ex post legal remedies if plaintiffs prove discrimination. Under many of these laws, such as equal protection or disparate

78. *See, e.g.*, Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265, 310 (1978) (expressing anxiety over harming innocent third parties); *see also supra* text accompanying note 24.

79. *See, e.g.*, Wygant v. Jackson Bd. of Educ., 476 U.S. 267, 274 (1986) (Powell, J.) (plurality opinion) ("[T]he role model theory . . . has no logical stopping point. The role model theory allows the Board to engage in discriminatory hiring and layoff practices long past the point required by any legitimate remedial purpose. Indeed, by tying the required percentage of minority teachers to the percentage of minority students, it requires just the sort of year-to-year calibration the Court stated was unnecessary . . . ."). To be sure, the Court recently upheld the narrowly tailored affirmative action program at the University of Michigan Law School. Grutter v. Bollinger, 539 U.S. 306, 343 (2003). But in doing so, the Court substantially altered what narrow tailoring had come to mean. For a sharp critique of how *Grutter* and *Gratz* confused the "narrow tailoring" requirements by replacing a "minimum necessary preference" requirement with a fuzzy "individualized consideration" requirement, see Ian Ayres & Sydney Foster, *Don't Tell, Don't Ask: Narrow Tailoring After Grutter and Gratz* (Yale Olin Paper No. 287, 2005), *available at* http://lsr.nellco.org/cgi/viewcontent.cgi?article=1030&context=yale/lepp.

80. *See generally* Linda Hamilton Krieger, *Civil Rights Perestroika: Intergroup Relations after Affirmative Action*, 86 CALIF. L. REV. 1251, 1276-1329 (1998); *see also* Deana A. Pollard, *Unconscious Bias and Self-Critical Analysis: The Case for a Qualified Evidentiary Equal Employment Opportunity Privilege*, 74 WASH. L. REV. 913, 926-37 (1999) (noting that prejudiced responses are largely unconscious, but antidiscrimination legislation requires a showing of intent to discriminate to obtain relief in most circumstances).

treatment, the ex ante command refers to intentional discrimination—purposefully different treatment of individuals because of their group membership. But such an explicit ex ante exhortation not to be intentionally unfair will do little to counter implicit cognitive processes, which take place outside our awareness yet influence our behavior.[81]

Ex post rights to sue have additional difficulties that can render them useless in the face of discrimination caused by implicit bias. Most obviously, they require the victim to perceive the discrimination. When the harm is invisible to the victim, talk of ex post remedies becomes moot. For various psychological reasons, invisibility runs deep. First, individuals tend to think that they are exceptional in that even though other members of their social category suffer from discrimination, they believe that they have gotten off relatively easy.[82] Second, when exposed to data on a case-by-case basis as compared to a big picture summary, individuals do poorly in spotting discrimination.[83] Third, through system justification motives, as explained by Gary Blasi and John Jost in this Symposium, the victim may see her fate as normal and deserved.[84] Fourth, even when a victim suspects discrimination, high transaction costs and difficult evidentiary burdens make litigation unlikely.[85]

A model that supposes that discrimination takes place explicitly, through a rational cost-benefit analysis or other expression of explicitly held views has become woefully out-of-date. A behavioral realist analysis has demonstrated that such a model of explicit discrimination is not up to

---

81. Michael Selmi has already made this point, drawing on an older body of scientific evidence of unconscious discrimination. *See* Michael Selmi, *Testing for Equality: Merit, Efficiency, and the Affirmative Action Debate*, 42 UCLA L. REV. 1251, 1283 (1995).

82. *See* Faye J. Crosby, *Understanding Affirmative Action*, 15 BASIC & APPLIED SOC. PSYCHOL. 13, 24-25 (1994) (reporting studies). This is sometimes called the "personal group discrimination dissociation" (PGDD).

83. *See* Krieger, *Civil Rights Perestroika supra* note 80, at 1305-09 (summarizing studies by Faye Crosby and Diane Cordova). Gary Blasi identifies other changes in the modern workplace, in which "shifting networks of contracting entities" replace traditional internal labor markets of large firms, which makes comparisons in treatment still more difficult. Gary Blasi, Default Discrimination: Dealing with Universal Bias Draft 3.0A 2005 (unpublished manuscript, on file with authors); *see also supra* text accompanying note 37; *see also* Samuel R. Bagenstos, *The Structural Turn and the Limits of Antidiscrimination Law*, 94 CALIF. L. REV. 1 (2006) (addressing difficulties of "innumerable daily encounters" in increasingly flat, flexible, boundaryless work arrangements).

84. *See* Gary Blasi and John T. Jost, *System Justification Theory and Research: Implications for Law, Legal Advocacy, and Social Justice*, 94 CALIF. L. REV. 1136-37 (2006).

85. *See, e.g.*, Michael J. Yelnosky, *Title VII, Mediation, and Collective Action*, 1999 U. ILL. L. REV. 583 (1999). Michael Yelonsky notes that "[w]hile approximately 80,000 charges are filed with the Equal Employment Opportunity Commission (EEOC) each year, many employees who believe their employer or prospective employer violated Title VII do not sue." *Id.* at 586. He attributes low filing rates to the fact that litigation requires plaintiffs to bear "financial, emotional, and reputational costs" in exchange for an uncertain chance of success due to the "rigid, highly stylized burdens of pleading and proof" of a Title VII claim. *Id.* at 588.

the task of responding to implicit bias, which is pervasive but diffuse, consequential but unintended, ubiquitous but invisible.[86]

To be fair, since *Griggs v. Duke Power Co.*[87] adopted a disparate impact theory of Title VII, antidiscrimination law has understood the problem of discrimination more capaciously.[88] And as the Krieger and Fiske article in this Symposium shows, even disparate treatment law could adapt to incorporate the new implicit cognitive learning.[89] We encourage such doctrinal evolution. But such projects should complement, not foreclose, a simultaneous exploration of other voluntary measures to prevent and remedy worse treatment actuated by implicit bias. We should not rigidly circumscribe fair measures to the status quo's anti-discrimination law.

Instead, we need a new model of discrimination for implicit bias—one based on a more accurate model of human cognition and emotion, especially its constraints. This new model must promote proactive structural interventions that minimize harm without relying solely on potential individual litigation.[90] A public health comparison is illuminating.[91] Public health is not pursued simply by creating ex post individual rights of action against those who intentionally "cause" disease. Instead, health agencies engage in preventative structural measures. For example, underlying clean water requirements is the notion that harmful agents, such as bacteria that an individual can spread to an entire community, are likely to go undetected by individual consumers and citizens. It is thus unreasonable to suppose that individuals alone, through conscious practice, will abate the problem. Rather, collective public health intervention is necessary. In fact, where water safety cannot be guaranteed, we do not wait until citizens get

---

86. *See* Bagenstos, *supra* note 83, at 9-12 ("Recognition of the pervasiveness of implicit bias lends support to a structural approach to antidiscrimination law.").

87. 401 U.S. 424, 432 (1971).

88. Many commentators have argued that disparate impact has only had a modest effect. *See, e.g.*, John J. Donohue III & Peter Siegelman, *The Changing Nature of Employment Discrimination Litigation*, 43 STAN. L. REV. 983, 998 (1991) (modest impact on litigation volume); Elaine W. Shoben, *Disparate Impact Theory in Employment Discrimination: What's* Griggs *Still Good For? What Not?*, 42 BRANDEIS L.J. 597, 597 (2004) (modest impact even after Civil Rights Act of 1991).

89. *See* Krieger & Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 CALIF. L. REV. 997 (2006).

90. *Cf.* Susan Sturm, *Second Generation Employment Discrimination: A Structural Approach*, 101 COLUM. L. REV. 458, 460-63 (2001) (proposing structuralism, which is "the development of institutions and processes to enact general norms in particular contexts," to combat second generation employment discrimination that may be result of "cognitive or unconscious bias"). Bagenstos is, however, pessimistic about the likelihood that a structural approach can be successfully implemented. *See* Bagenstos, *supra* note 83. His criticisms apply, however, more forcefully to ex post lawsuits than to ex ante, voluntary adoption of fair measures.

91. A comparison between environmental and civil rights law is also illuminating. *See* Tseming Yang, *The Form and Substance of Environmental Justice: The Challenge of Title VI of the Civil Rights Act of 1964 for Environmental Regulation*, 29 B.C. ENVTL. AFF. L. REV. 143 (2002). Yang points out that civil rights law "ignores the fact that discrimination, much like environmental degradation, is an aspect of life that is pervasive throughout society . . . ." *Id.* at 195.

infected; instead, we inject a purifying agent prior to imbibing. We are willing to take these preventative, proactive measures partly because we recognize that these problems cannot be easily detected by individuals, produce demonstrable harm, and reflect present concerns, not mere sediments of some distant, eccentric, pathological past. By contrast, we respond differently to a truly historical problem, such as smallpox, which has been eradicated, by lowering our guard and devoting minimal resources to detecting recurrences.

## II
## THE PSYCHOLOGICAL CONSTRUCTION OF MERIT

### *A. Conventional Thinking: Sacrificing Merit*

Opponents decry affirmative action as a deviation from merit. Skeptical about the degree of discrimination that persists, they see underrepresentation of women and minorities as the real-world consequences of actual merit differentials. Some opponents view these differences as simply the state of the world, either freely chosen,[92] genetically predetermined, or the end-result of a beneficial Social Darwinism. Other opponents seem more troubled by the differences that result from group-based disadvantage, but this concern does not alter their view that affirmative action sacrifices merit.

Proponents of affirmative action can adopt one of three standard responses, which we label as (a) net benefit, (b) merit as fraud, and (c) institutional mission. First, and most conservative, *net benefit* concedes that affirmative action sacrifices merit but suggests that the social justice and social stability benefits of affirmative action outweigh the efficiency costs.[93] This proponent of affirmative action weights the benefits and costs differently from a utilitarian opponent of affirmative action. Of course, opponents of affirmative action who view colorblindness and/or selection by merit as moral or constitutional imperatives claim to be unwilling to engage in such policy trade-offs.[94]

---

92. One might view a career choice that accurately reflects a person's preferences to be freely chosen. But what if one's preferences are influenced by cultural stereotypes? For an ISC explication of this interrelationship, see Brian A. Nosek et al., *Math = Male, Me = Female, Therefore Math ≠ Me*, 83 J. PERSONALITY & SOC. PSYCHOL. 44 (2002). The authors demonstrate how background cultural stereotypes that math is not a female strength discourages women from wanting to study math.

93. *See, e.g.*, Richard H. Fallon, *Affirmative Action Based on Economic Disadvantage*, 43 UCLA L. REV. 1913, 1930 (1996) (describing a "non-merit-based" form of affirmative action, which would allow for some sacrifice of traditional ideas of merit for other benefits); Jerry Kang, *Negative Action Against Asian American: The Internal Instability of Dworkin's Defense of Affirmative Action*, 31 HARV. C.R.-C.L. L. REV 1, 6 (1996) (discussing Ronald Dworkin's defense of affirmative action, which supposes a "net benefit" condition).

94. We say "claim to be," because these very same individuals will often accept color consciousness when it comes to racial profiling in a post 9/11 world. *See, e.g.*, Jerry Kang, *Thinking*

Second, and least conservative, *merit as fraud* challenges prevailing merit definitions as fundamentally biased.[95] On this view, for example, standardized tests do not examine for anything resembling intelligence or aptitude; rather, they merely reify past privilege.[96] Those with the most resources determine the nature of such tests to keep power within traditionally privileged circles.[97] Even if such an effort is not conscious, it nonetheless emphasizes a self-privileging view of merit.

Third, situated between these two extremes, *institutional mission* emphasizes the relational nature of merit: what counts as merit depends on the goal.[98] A brilliant mathematical ability is not merit if the goal is to win a full-contact cage match. This critique recasts the debate on merit as a debate on institutional mission. For an institution of higher education, is the goal to admit the most intelligent as defined as the best test takers? Or is its mission broader, for example, including the goal of training future leaders? If it includes the latter, then a university must seek "merit" in evidence likely to predict a future leader, even at the expense of standardized test scores or grades.

The rhetorical back-and-forth between these various positions on merit is well rehearsed. What does the implicit social cognition (ISC) have to add to this debate?

### *B. Behavioral Realism: Mismeasuring Merit*

It is tempting to pursue a behavioral realist critique that provides evidence for and explores the implications of the *merit as fraud* or *institutional mission* positions. But we want to confront the hardest case for affirmative action by accepting the status quo conception of merit. Although we have serious reservations about this conception,[99] we put

---

*Through Internment: 12/7 and 9/11*, 9 ASIAN L.J. 195, 200 (2002); Leti Volpp, *The Citizen and the Terrorist*, 49 UCLA L. REV. 1575, 1576-77 (2002).

95. *See, e.g.*, JOAN WILLIAMS, UNBENDING GENDER: WHY FAMILY AND WORK CONFLICT AND WHAT TO DO ABOUT IT 213-17 (2000) (discussing how what counts as meritorious is designed around masculine norms); Richard Delgado, *Official Elitism or Institutional Self Interest? 10 Reasons Why UC-Davis Should Abandon the LSAT (And Why Other Good Law Schools Should Follow Suit)*, 34 U.C. DAVIS L. REV. 593 (2001).

96. *See, e.g.*, Roithmayr, *supra* note 64, at 734.

97. *See, e.g.*, Daria Roithmayr, *Deconstructing the Distinction Between Bias and Merit*, 85 CALIF. L. REV. 1449 (1997).

98. *See, e.g.*, Kang, *supra* note 93, at 8; Susan Sturm & Lani Guinier, *The Future Of Affirmative Action: Reclaiming The Innovative Ideal*, 84 CALIF. L. REV. 953, 968-69 (1996) (embracing the idea of "functional merit" rather than merit as a concept of desert); Kenneth L. Karst & Harold W. Horowitz, *Affirmative Action and Equal Protection*, 60 VA. L. REV. 955, 965 (1974) (defining merit as that which satisfies social needs).

99. *See, e.g.*, Faye J. Crosby, et al., *Affirmative Action: Psychological Data and the Policy Debates*, 58 AM. PSYCHOLOGIST. 93, 100 (2003) (pointing out how weakly the SAT predicts performance).

them aside and probe what ISC has to say about how we conventionally measure merit.

### *1. Perceiver Effects*

The mind does its work silently. One does not hear whirring as thought processes change gear, or the sound of draining as information is lost. Nor does one receive a printout of errors at the end of the day. But the human attribute of self-consciousness, which includes the ability to reflect on the contents of one's own mind, gives the illusion of access and control.[100] "I know what I know, I know what I believe, I can change what I think"—these may be true of our self-consciously endorsed attitudes and beliefs, but this reflects only a fraction of the work our minds do.

Implicit cognitive processes influence how we, as perceivers, judge others. There is now overwhelming evidence that mental constructs that are cognitively accessible influence how the perceiver evaluates and judges others. The standard experiments evincing this phenomenon trigger a particular construct, then require the participant to evaluate some person's action. For example, researchers ask participants to read passages designed to activate particular personality qualities, such as "stubborn" or "persistent," then instruct them to evaluate ambiguous target behavior. When participants do so, their evaluations are biased in accordance with the activated knowledge.[101] This phenomenon has been demonstrated over myriad domains, such as: wanting to work with a gay person;[102] judging alcoholics;[103] interpreting aggressive behavior;[104] and treating women in sexist ways.[105]

Significantly, the mental constructs that guide our evaluations include beliefs (stereotypes) and feelings (prejudice) about entire social categories.

---

100. *See generally* DANIEL M. WEGNER, ILLUSION OF CONSCIOUS WILL (2003).

101. *See* Constantine Sedikides & John J. Skowronski, *Towards Reconciling Personality and Social Psychology: A Construct Accessibility Approach*, 5 J. SOC. BEHAVIOR & PERSONALITY 531, 534-36 (1990) (discussing studies that show that subjects have a bias in favor of interpreting ambiguous target behavior in accordance with information (or constructs) accessible, both chronically (as with stereotypes) or momentarily (as primed and activated by researchers)).

102. *See* James Johnson et al., *Construct Accessibility, AIDS, and Judgment*, 9 J. SOC. BEHAV. & PERSONALITY 191, 195-98 (1994) (demonstrating that subjects primed with information negatively associating gays with AIDS reported a lower desire to work with the target, a gay job applicant).

103. *See* Lillian Southwick, Claude Steele & Michael Lindell, *The Roles of Historical Experience and Construct Accessibility in Judgments About Alcoholism*, 10 COGNITIVE THERAPY & RES. 167, 182 (1986) (showing that an "alcoholic" prime prompted more construct-consistent judgments than actual familial or friendship experience with an alcoholic).

104. *See* Sandra Graham & Cynthia Hudley, *Attributions of Aggressive and Nonaggressive African-American Male Early Adolescents: A Study of Construct Accessibility*, 30 DEV. PSYCHOL. 365, 369-71 (1994) (showing that for non-aggressive children, priming to perceive negative events as intentionally caused produces more extreme responses).

105. *See* Laurie A. Rudman & Eugene Borgida, *The Afterglow of Construct Accessibility: The Behavioral Consequences of Priming Men to View Women as Sexual Objects*, 31 J. EXPERIMENTAL SOC. PSYCHOL. 493, 511-13 (1995) (priming perceivers to categorize women as sexual objects resulted in more sexist behavior toward female targets compared to unprimed perceivers).

Unconscious stereotypes, rooted in social categorization, are ubiquitous and chronically accessible.[106] They are automatically prompted by the mere presence of a target mapped into a particular social category.[107] Thus, when we see a Black (or a White) person, the attitude and stereotypes associated with that racial category automatically activate. Further, these attitudes and stereotypes influence our judgments,[108] as well as inhibit countertypical associations.[109]

---

106. *See* Susan T. Fiske, *Stereotyping, Prejudice, and Discrimination, in* 2 THE HANDBOOK OF SOCIAL PSYCHOLOGY 357, 364 (D.T. Gilbert, S.T. Fiske & G. Lindzey eds., 1998) (discussing studies showing that prejudice and stereotypes can be learned and operate unconsciously).

107. *See* Mahzarin R. Banaji & Curtis Hardin, *Automatic Stereotyping*, 7 PSYCHOL. SCI. 136, 140-41 (1996) (showing gender information primed through words produced faster (automatic) responses to targets with stereotypical gender roles (e.g., doctor-he) than counterstereotypic prime-target gender pairs (e.g., nurse-he)); Irene V. Blair & Mahzarin R. Banaji, *Automatic and Controlled Processes in Stereotype Priming*, 70 J. PERSONALITY & SOC. PSYCHOL. 1142, 1147-48 (1996) (showing that gender stereotype priming resulted in faster target responses to gender-stereotypic prime-target pairs (e.g., gentle-Jane) than counterstereotypic trials (e.g., strong-Jane), arguing that stereotypes may be automatically activated through a priming procedure); Susan T. Fiske & Steven L. Neuberg, *A Continuum of Impression Formation, From Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation, in* 23 ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY 1, 4, 23-24 (M.P. Zanna ed., 1990) (discussing studies showing that initial categorization occurs immediately upon receiving information relevant to a meaningful social category (e.g., cognitive stereotyping)); Thomas E. Ford et al., *Influence of Social Category Accessibility and Category-Associated Trait Accessibility on Judgments of Individuals*, 12 SOC. COGNITION 149, 163-164 (1994) (showing that priming a narrowly defined social category activates judgments consistent with the primed category); David L. Hamilton & Jeffrey W. Sherman, *Stereotypes, in* 2 HANDBOOK OF SOCIAL COGNITION 1, 40-42 (R.S. Wyer, Jr. & T.K. Srull eds., 1994) (discussing studies demonstrating the automaticity of stereotyping); Charles W. Perdue et al., *"Us" and "Them": Social Categorization and the Process of Intergroup Bias*, 59 J. PERSONALITY & SOC. PSYCHOL. 475, 478-79, 482-84 (1990) (showing the use of words describing the in-group (e.g., us) resulted in faster and positive affective associations to unfamiliar targets, compared to words describing out-group status (e.g., them)).

108. *See* John F. Dovidio et al., *Racial Stereotypes: The Contents of Their Cognitive Representations*, 22 J. EXPERIMENTAL SOC. PSYCHOL. 22, 32-33 (1986) (showing that participants responded faster to the activation of traits stereotypic of the prime category (e.g., White persons are ambitious) compared to traits counterstereotypic of the prime category (e.g., White persons are musical)); Jack A. Glaser & Mahzarin R. Banaji, *When Fair Is Foul and Foul Is Fair: Reverse Priming in Automatic Evaluation*, 77 J. PERSONALITY & SOC. PSYCHOL. 669, 671 (1999) (showing when race-neutral primes were used, congruent prime-targets pairs (e.g., negative-black and positive-white) exhibited slower results than incongruent pairs (e.g., positive-black and negative-white)); Lorella Lepore & Rupert Brown, *Category and Stereotype Activation: Is Prejudice Inevitable?*, 72 J. PERSONALITY & SOC. PSYCHOL. 275, 281-82 (1997) (showing that despite common stereotype knowledge, high-prejudiced participants formed a more negative impression of the target person after subliminal priming of the category "Blacks" than participants in the non-condition, while low-prejudiced participants formed the opposite impressions); William Von Hippel et al., *The Linguistic Intergroup Bias As an Implicit Indicator of Prejudice*, 33 J. EXPERIMENTAL SOC. PSYCHOL. 490, 507 (1997) (showing that a subject's implicit prejudice towards African-Americans resulted in differential responses to African-American and Caucasian targets).

109. *See* Ad van Knippenberg & Ap Dijksterhuis, *A Posteriori Stereotype Activation: The Preservation of Stereotypes Through Memory Distortion*, 14 SOC. COGNITION 21, 46-48 (1996) (showing that stereotype activation inhibits and weakens the ability of perceivers to recall stereotype-inconsistent behavior); Yaacov Trope & Erik P. Thompson, *Looking For Truth in All the Wrong Places? Asymmetric Search of Individuating Information About Stereotyped Group Members*, 73 J.

As applied to race, Jerry Kang has labeled this process as a sort of "racial mechanics."[110] An individual (target) is mapped into a social category in accordance with prevailing legal and cultural mapping rules. Once mapped, the category activates various meanings, which include cognitive and affective associations that may be partly hard-wired but are mostly culturally-conditioned. These activated meanings then alter the interaction between perceiver and target. These mechanics occur automatically, without effort or conscious awareness on the part of the perceiver.[111] Although perceivers assume that their judgments are based "on the merits"— in other words on the basis of qualities that the target in fact exhibits—the truth is more complicated. Even if we lack animus, intention to discriminate, or self-awareness of bias, our judgments of others may still lack "mental due process."[112] On subjective measures of merit, the perceiver's (evaluator's) expectations guide what she actually sees in the target (the person being evaluated). In more plain language, if we expect someone to be violent, we will likely see violence when presented with ambiguous behavior.[113]

From the vantage point of social psychology, these cognitive processes are old news. Even in the law reviews, Linda Hamilton Krieger set

---

PERSONALITY & SOC. PSYCHOL. 229, 235 (1997) (showing that participants asked fewer individuating questions of stereotyped targets than non-stereotyped targets).

110. *See generally* Kang, *supra* note 12, at 1497-1506. At least as applied to race, these mechanics are largely socially constructed. In other words, the recognition of particular races, the legal and cultural rules by which we map individual human beings into racial categories, and the meanings (both attitudes and stereotypes) associated with these categories are all principally products of human culture and institutions. *See id.* at 1501-02.

111. Scores of studies demonstrate that subliminal priming can alter the ways we interpret ambiguous behavior. *See, e.g.*, John A. Bargh et al., *Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action*, 71 J. PERSONALITY & SOC. PSYCHOL. 230, 236-38 (1996) (demonstrating that indirect exposure to words associated with the elderly altered the speed of walking down a hallway); Patricia G. Devine, *Stereotypes and Prejudice: Their Automatic and Controlled Components*, 56 J. PERSONALITY & SOC. PSYCHOL. 5, 11-12 (1989) (demonstrating that subliminal priming with words stereotypically associated with Blacks can cause perceivers to evaluate ambiguous behavior as more "aggressive"); John F. Dovidio et al., *On the Nature of Prejudice: Automatic and Controlled Processes*, 33 J. EXPERIMENTAL SOC. PSYCHOL. 510, 516-17 (1997) (demonstrating that subliminal flashes of Black or White faces can produce time differentials in classifying positive or negative words).

112. Mahzarin R. Banaji & R. Bhaskar, *Implicit Stereotypes and Memory: The Bounded Rationality of Social Beliefs*, *in* MEMORY, BRAIN, AND BELIEF 139-175 (D. L. Schacter & E. Scarry, eds., 2000).

113. *See* Birt L. Duncan, *Differential Social Perception and Attribution of Intergroup Violence: Testing the Lower Limits of Stereotyping of Blacks*, 34 J. PERSONALITY & SOC. PSYCHOL. 590 (1976) (showing that the race of a "shover" in a video altered whether a "shove" was deemed aggressive). If the shover was Black and the victim was White, 75% of the perceivers characterized it as aggressive; by contrast, if the shover was White and the victim was Black, only 17% of the perceivers thought it aggressive. *Id.* at 595. *See also* H. Andrew Sagar & Janet Ward Schofield, *Racial and Behavioral Cues in Black and White Children's Perceptions of Ambiguously Aggressive Acts*, 39 J. PERSONALITY & SOC. PSYCHOL. 590, 593-95 (1980) (showing that the darkness of the skin of drawn characters altered whether a hallway bump in an ambiguous narrative was viewed as hostile by both White and Black participants).

out much of this science back in 1995[114] and 1998.[115] What is new, however, is an updating of the empirical evidence for such biased evaluation. We have already described the evidence of predictive validity, as provided by the meta-analysis by Poehlman, et al.[116] Many of the studies included in that analysis addressed behavioral consequences in judging or evaluating others' merit.[117] For example, implicit bias as measured by the IAT has been correlated with biased evaluations of job candidates. Laurie Rudman and Peter Glick have demonstrated that negative evaluations of agentic (self-promoting, highly competent) women relative to identically characterized men correlated with IAT scores but not to explicit self-reports about belief in gender stereotypes.[118]

### 2. *Target Effects*

It is fairly easy to see how subjective evaluations can be biased, but what about completely objective measures, such as standardized multiple choice tests? Numerous commentators have challenged these tests as insufficiently validated; in other words, there is little evidence that the tests actually "test" for the right set of characteristics that are functionally necessary for a particular task.[119] But we tackle a harder question.

---

114. *See* Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161 (1995) (arguing that a large number of biased employment decisions result from a variety of unintentional categorization-related judgment errors characterizing normal human cognitive functioning).

115. *See, e.g.*, Kreiger, *supra* note 80, at 1268-69 (describing psychological processes, such as selective memory and expectancy-confirmation that could lead to schema-consistent interpretations).

116. Poehlman et al., *supra* note 46; *see also* Dolly Chugh, *Societal and Managerial Implications of Implicit Social Cognition: Why Milliseconds Matter*, 17 SOC. JUSTICE RES. 203 (2004) (reviewing scientific evidence of predictive validity relevant to the work that managers do). Chugh emphasizes that managers work in a frenetic mode, with half of their activities lasting less than nine minutes and ninety-three percent of their verbal contact being ad hoc, not pre-planned. *Id.* at 205. If time, concentration, and effort are necessary to prevent implicit bias from influencing behaviors, managers do not seem to have such resources aplenty in their daily worklife.

117. Back in 1999, Amy Wax argued that the data were simply not available on type of bias, magnitude, and consequence. *See* Amy L. Wax, *Discrimination as Accident*, 74 IND. L.J. 1129 (1999). Although that may have been true then, it is no longer the case now. For a trenchant, contemporaneous critique to Wax's paper, see Michael Selmi, *Response to Professor Wax, Discrimination as Accident: Old Whine, New Bottle*, 74 IND. L.J. 1233 (1999).

118. *See* Rudman & Glick, *supra* note 56, at 747-48.

119. *See, e.g.*, Sturm & Guinier, *supra* note 98, at 969-80. Robert Sternberg has identified three components to "successful intelligence": the ability to think "analytically, creatively, and practically." ROBERT J. STERNBERG, SUCCESSFUL INTELLIGENCE 127 (1996). Because "only analytical intelligence is valued on tests and in the classroom" such measures are often inaccurate indicators of how well one will perform in advance schooling or in a career. *Id.* at 127, 127-37; *see also* Howard Gardner, *Cracking Open the IQ Box*, *in* THE BELL CURVE WARS 23, 29 (Steven Fraser ed., 1995) (noting the increase in "performance examinations" because standardized tests focus solely on important, but overly narrow aspects of intelligence); Robert J. Sternberg & Wendy M. Williams, *Does the Graduate Record Examination Predict Meaningful Success in the Graduate Training of Psychologists?*, 52 AM. PSYCHOLOGIST. 630, 636-37 (1997) (demonstrating that GRE test scores modestly predicted first-year grades, but failed to predict second-year grades, and that the GRE Analytical test score predicted the

Suppose that a test does in fact measure the correct characteristics. Even so, implicit cognitive processes within the test-taker can produce differences in test performance, as a function of arbitrary environmental cues. They can do so in part by altering how the perceiver thinks about herself, which can substantially hamper (and sometimes improve) performance.[120] Such studies raise fundamental questions. To what extent is the measure predictive if it "moves" with trivial interventions such as reminding people of their social group?[121] We focus on the "stereotype threat" literature, which has received serious attention from scientists, the public, and even the Educational Testing Service[122] whose very existence rests on the general public's confidence in its standardized tests.

Individuals who belong to social groups marked by negative stereotypes about intellectual performance underperform when cues remind them of their group identity. In their seminal experiment, Claude Steele and Joshua Aronson gave a difficult verbal test to White and Black undergraduate students. One group was told that the test measured how smart they were. Another comparable group was told that the (identical) test was simply a laboratory exercise. In the latter condition, the Black students performed as well as the White students, controlling for the participants' initial skills. But in the former condition, Black students greatly underperformed equally skilled White students.[123] As the authors explain: "[T]he existence of a negative stereotype about a group to which one

---

"ratings of analytical, creative, practical, research, and teaching abilities by primary advisers and ratings of dissertation quality by faculty readers" for men but not for women). Many scholars and scientists challenged the validity of standardized testing in response to the controversial book, the THE BELL CURVE. *See, e.g.*, THE BELL CURVE DEBATE (Russell Jacoby & Naomi Glauberman eds., 1995); CLAUDE S. FISCHER ET. AL., INEQUALITY BY DESIGN: CRACKING THE BELL CURVE MYTH (1996); INTELLIGENCE, GENES, AND SUCCESS: SCIENTISTS RESPOND TO THE BELL CURVE (Bernie Devlin et. al. eds., 1997); MEASURED LIES: THE BELL CURVE EXAMINED (Joe L. Kincheloe et. al. eds., 1996); Gardner, *supra*.

120. For example, unconscious activation of one's significant other who is either critical or accepting can prompt consistent self-evaluations. *See* Mark W. Baldwin, *Primed Relational Schemas as a Source of Self-Evaluative Reactions*, 13 J. SOC. & CLINICAL PSYCHOL. 380 (1994); *see also* Mark W. Baldwin et. al., *Priming Relationship Schemas: My Advisor and the Pope are Watching Me from the Back of My Mind*, 26 J. EXPERIMENTAL SOC. PSYCHOL. 435 (1990) (showing self-evaluation consistent with approving or disapproving subliminal primes of "personally significant authority figures").

121. One could ask the same question about even more substantial interventions, such as test preparation courses. In one of our labs, a student recently improved his GRE score by a good 250 points by cramming for a month prior to the second test, making a crucial difference between getting into a mediocre graduate program and a highly selective one. The fact that a mere month's worth of study can radically change this measure suggests that we be cautious about its interpretation.

122. *See, e.g.*, ALYSSA M. WALTERS ET AL., EDUCATIONAL TESTING SERVICE, STEREOTYPE THREAT, THE TEST-CENTER ENVIRONMENT, AND PERFORMANCE ON THE GRE GENERAL TEST 34-36 (2004), *available at* http://ftp.ets.org/pub/gre/gre-01-03R.pdf (studying environmental cues that might trigger stereotype threat in GRE testing centers).

123. *See* Claude M. Steele, *A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance*, 52 AM. PSYCHOLOGIST. 613, 620 (1997). "Analysis of covariance was used to remove the influence of participants' initial skills, measured by their verbal SAT scores, on their test performance." *Id.*

belongs . . . means that in situations where the stereotype is applicable, one is at risk of confirming it as a self-characterization, both to one's self and to others who know the stereotype. This is what is meant by stereotype threat."[124] This possible confirmation triggered an anxiety that somehow disrupted performance. Although the precise mechanism of this phenomenon is still not well understood,[125] there are now incontrovertible data of such performance-disruption. These experiments include Blacks underperforming on tests of intellectual ability;[126] women underperforming on tests of mathematical ability; elderly underperforming on memory tests;[127] and low socio-economic status students underperforming on verbal ability tests.[128] In certain contexts, Whites are also subject to stereotype-threat, for example when reminded of Asian superiority before taking a math exam.[129] Researchers have replicated these results across a broad developmental span, ranging from elementary school to college.[130] Not only are intelligence tests vulnerable; researchers have also found evidence of stereotype-threat in sporting activities.[131]

---

124. Claude M. Steele & Joshua Aronson, *Stereotype Threat and the Intellectual Test Performance of African Americans*, 69 J. PERSONALITY & SOC. PSYCHOL. 797, 808 (1995).

125. *See* Jessi L. Smith, *Understanding the Process of Stereotype Threat: A Review of Mediational Variables and New Performance Goal Directions*, 16 EDUC. PSYCHOL. REV. 177, 178 (2004) (providing the inconclusive results of various stereotype threat mechanism studies, which considered anxiety, evaluation apprehension, performance confidence, effort, self-handicapping, perceptions of a test fairness, stereotype endorsement, and other individual differences); *see also* S. Christian Wheeler & Richard E. Petty, *The Effects Of Stereotype Activation On Behavior: A Review Of Possible Mechanisms*, 127 PSYCHOL. BULL. 797 (2001) (discussing mechanisms).

126. *See generally* Steele, *supra* note 123.

127. Becca Levy, *Improving Memory in Old Age Through Implicit Self-Stereotyping*, 71 J. PERSONALITY & SOC. PSYCHOL. 1092, 1092-1101 (finding subliminal priming of positive stereotypes of aging improved memory performance for elderly subjects while priming of negative stereotypes of aging worsened memory performance).

128. Jean-Claude Croizet & Theresa Claire, *Extending the Concept of Stereotype Threat to Social Class: The Intellectual Underperformance of Students from Low Socioeconomic Backgrounds*, 24 PERSONALITY & SOC. PSYCHOL. BULL. 588 (1998) (demonstrating that when SES stereotypes were triggered by asking students about their "parents' occupation and education level," low SES students performed worse than high SES students when a verbal test was presented as a measure of "intellectual ability" but not when presented as a memory test).

129. *See* Joshua Aronson, et al., *When White Men Can't Do Math: Necessary and Sufficient Factors in Stereotype Threat*, 35 J. EXPERIMENTAL SOC. PSYCHOL. 29, 38 (1999). White (including Jewish) students who had an average SAT math score of 712.17 took a difficult eighteen question GRE math test. Those under the stereotype threat condition were exposed to newspaper articles about Asian superiority in math; the control group was not. Performance was substantially depressed by the stereotype threat (number of accurate answers $M = 6.55$ versus $M = 9.58$ for the control group; $p < .01$). *See id.* at 33-34.

130. *See, e.g.*, N. Ambady et al., *Stereotype Susceptibility in Children: Effects of Identity Activation on Quantitative Performance*, 12 PSYCHOL. SCI. 385 (2001); C. McKown & R. S. Weinstein, *The Development and Consequences Of Stereotype Consciousness In Middle Childhood*, 74 CHILD DEV. 498 (2003).

131. J. Stone et al., *Stereotype Threat Effects on Black and White Athletic Performance*, 77 J. PERSONALITY & SOC. PSYCHOL. 1213 (1999).

Researchers have also found evidence of two additional effects: stereotype boost[132] and stereotype lift.[133] Stereotype-*boost* takes place when unconscious activation of a particular identity improves, not depresses, performance.[134] Margaret Shih and her colleagues first demonstrated this phenomenon by testing Asian American women on difficult math tests, after subliminally priming them with an Asian, female, or control category. They found that the participant group primed with the Asian identity performed best (an accuracy of $M = 54\%$ on a difficult twelve question math test), the control group that had no identity primed came next, and the group primed with its female identity performed the worst (accuracy of $M = 43\%$).[135] Thus the "Asian American" cue raised math performance, whereas the "female" cue decreased math performance.

A related phenomenon is stereotype *lift*, which is increased performance "caused by the awareness that an outgroup is negatively stereotyped."[136] This finding of lift is more insidious than the finding of boost because it shows that the derogation of outgroups can improve one's own scores. Through a meta-analysis of stereotype threat studies, Gregory Walton and Geoffrey Cohen focused solely on the performance of White men. Specifically, they compared how White men performed in conditions designed to trigger stereotype-threat in others as compared to how White men performed in the control conditions. They found that White men performed better in the former condition.[137] In terms of effect sizes, translated to the SAT scale, White males received a fifty-point advantage.[138]

These are surprising and disturbing findings. Such effects do not explain entirely the differentials in testing across various social categories. But they should give us pause as we confront the fact that arbitrary environmental cues can trigger implicit cognitive processes that interfere with or facilitate performance on seemingly objective measures. What we thought to be fair assessments of "merit" can turn out to be mismeasurements—not because of explicit animus but because of hidden mental processes that by their nature cannot reach conscious awareness.[139]

---

132. *See* Margaret Shih et al., *Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance*, 10 PSYCHOL. SCI. 80 (1999) [hereinafter *Stereotype Susceptibility*]; Margaret Shih et al., *Stereotype Performance Boosts: The Impact of Self-Relevance and the Manner of Stereotype Activation*, 83 J. PERSONALITY & SOC. PSYCHOL. 638, 638 (2002) (discussing studies).

133. *See* Gregory M. Walton & Geoffrey L. Cohen, *Stereotype Lift*, 39 J. EXPERIMENTAL SOC. PSYCHOL. 456 (2003).

134. Shih, *Stereotype Susceptibility*, *supra* note 132, at 80-81.

135. *See id.* at 81. These differences were statistically significant under a linear contrast analysis ($p < .05$).

136. *See* Walton & Cohen, *supra* note 133, at 456.

137. *Id.* at 463 ($d = 0.24$; $p < 0.0001$).

138. *See id.*

139. One could argue that there is in fact no mismeasurement because those who suffer from permanently debilitating performance drags, regardless of their unfortunate causes, have less merit. At bottom, this is a definitional claim about what "merit" encompasses. But notice that the disruptions are

In sum, as perceivers, we may misperceive, even though we honestly believe we are fair and just. As targets, we may underperform, even though we proudly assert immunity from negative stereotypes about our identity groups. These mismeasurements have immediate consequences that can extend into the future, by creating self-fulfilling prophecies that generate long-term path dependencies. Worse, these errors are not randomly dispersed and hence likely to wash out over time; instead, they have a systematic tilt in the direction of the implicit bias. As discussed in Part I, problems in the future will not be easy to remedy on the basis of unfairness experienced in the past. Accordingly, we have even more reason to root out mismeasurements of merit now.

### C. *Better Measures of Merit*

If current measures of merit are defective, and we have reasons to be wary of both subjective and objective measures, what is to be done? We proffer no silver bullets, for the science provides none. Instead, we provide a few modest interventions that address both perceiver and target effects. Given space constraints, we only sketch out suggestions, which we hope will be pursued in greater detail as part of a behavioral realist research agenda.

#### 1. *Motivate Decision Makers to Correct Bias by Increasing Self Awareness*

As a threshold matter, in order to correct bias, decision makers in admissions, hiring, and contracting must be made aware of their own implicit biases. Since so many of us are convinced that we are race- or gender-blind, we tend to dismiss evidence of pervasive implicit bias as somehow inapplicable to ourselves.[140] In other words, we assume that we are

---

not always stable and inevitable; rather, they are often highly reactive to small changes in environmental conditions. Because merit traditionally understood reflects stable characteristics internal to the individual, it seems odd now to incorporate erratic, environmentally induced disruptions into the definition. Further, on both efficiency and fairness grounds, it makes sense to prevent these disruptions instead of normalizing them as simply a part of merit.

140. Our call for increased self-awareness should not be misunderstood as a naïve embrace of "diversity training," regardless of its form. Professional consultants encourage self-awareness through various strategies. As numerous commentators have noted, such programs may or may not be successful along various metrics. *See, e.g.*, Kimberly D. Krawiec, *Cosmetic Compliance and the Failure of Negotiated Governance*, 81 WASH. U. L.Q. 487, 515 (2003) (noting some studies provide "little empirical support . . . that diversity training contributes to attitudinal or behavioral changes," and some evidence suggests such training decreases tolerance, but other studies indicate participants have "generally positive reactions to diversity training"); David B. Wilkins & G. Mitu Gulati, *Why Are There So Few Black Lawyers in Corporate Law Firms? An Institutional Analysis*, 84 CALIF. L. REV. 493, 593-95 (1996). Although we advocate self-awareness of implicit bias, we take no general stance on the larger dispute over diversity training.

somehow exceptional and immune from the cognitive errors that others make.[141] Accordingly, actual self-diagnosis should be encouraged.

In practical terms, this means that those who admit, hire, select, and evaluate should volunteer to experience their bias directly. Implementation costs are minimal because tests like the implicit association test (IAT) can be taken online, free of charge.[142] Numerous anecdotal reports suggest that the experience of the test creates a new form of self-awareness that is striking and persuasive.[143] Of course, some individuals may see little or no preference. But this too is valuable self-discovery. Among those who see an associational preference, many will protest that the test means nothing, which raises again the question of predictive validity of real-world discrimination. But by this point, any claim of total color- or gender-blindness is disproved; rather, the claim has shifted to behavioral neutrality notwithstanding mental preference. Even this increase in self-understanding is valuable because it motivates individuals to consider implementing personal and institutional processes that prevent behavioral manifestations of implicit bias. Finally, to the extent that fear of legal liability would discourage such self-testing, we agree with Deana Pollard that an evidentiary privilege should be carefully crafted and recognized.[144]

A call for increased self-awareness is neither new nor restricted to arguments based on ISC. For example, Susan Sturm has highlighted how Deloitte and Touche addressed the question of gender disparities in their business.[145] Until that firm actually measured the gender distribution of work assignments, it was unaware of how its informal procedures systematically doled out less desirable work to women.[146] Upon becoming aware of this issue, Deloitte and Touche instituted reforms, which included an

141. *See generally* David Alain Armor, The Illusion of Objectivity: A Bias in the Perception of Freedom From Bias (1998) (unpublished dissertation); Nandita Murukutla & David A. Armor, Illusions of Objectivity and the Dispute over Kashmir: An Experimental Test of the Effects of Disagreement (Oct. 7, 2004) (unpublished manuscript). Ironically, those who seem most confident of their objectivity may turn out to discriminate the most. *See* Eric Luis Uhlmann & Geoffrey L. Cohen, *Constructed Criteria: Redefining Merit to Justify Discrimination*, 16 PSYCHOL. SCI. 474, 479 (2005).

142. *See* Project Implicit, http://implicit.harvard.edu/implicit (last visited Jan. 14, 2006).

143. *See* Anthony G. Greenwald et al., *Consequential Validity of the Implicit Association Test: Comment on the Article by Blanton and Jaccard*, 61 AM. PSYCHOLOGIST. 56-61 (2006). (describing how the IAT produces a "palpable" experience of bias and calling this phenomenon potentially "its central asset"); Shankar Vedantam, *No Bias*, WASH. POST, Jan. 23, 2005, at W12 (providing anecdotes).

144. *See* Pollard, *supra* note 80, at 997-1018.

145. *See* Susan Sturm, *Second Generation Employment Discrimination: A Structural Approach*, 101 COLUM. L. REV. 458, 492-93 (2001) (pointing out that despite hiring women at a 50% rate, Deloitte's "rate of promotion hovered at around 10%").

146. *See id.* at 496 ("The Task Force found that on the accounting side, women's assignments tended to be clustered in not-for-profit companies, health care, and retail. . . . Women were rarely assigned to such high-potential areas as mergers and acquisitions.").

annual audit of work assignments.[147] This example demonstrates how we can achieve self-awareness by measuring actual outcomes (such as work distribution or rates of promotion) and comparing them to some baseline expectation. The IAT provides self-insight by measuring our associational preferences and comparing them to the baseline expectation of neutrality.

### 2. *Prevent the Influence of Implicit Bias*

#### a. *Cloak Social Category*

If an individual cannot be mapped to a racial or gender category because such information is successfully cloaked,[148] then implicit (or even explicit) bias cannot readily influence the evaluation. Thus, where feasible, we recommend cloaking social category in order to prevent biased perceptions.

Impressive evidence of the benefits of cloaking comes from the well-publicized studies of orchestra auditions. When musicians perform behind a screen, so that judges hear only the music and cannot see the performer, judges choose different musicians. In the early 1970s with pressure from unions, American Symphony Orchestras implemented blind auditioning. Consequently, more female musicians who played a variety of instruments joined the lone female harp.[149] The more recent "resume study" conducted by Marianne Bertrand and Sendhil Mullainathan provides further support for cloaking social categories. The researchers sent out fictional resumes that differed only with respect to whether the applicant had a Black- or White-sounding name (for example, Jamaal Jones vs. James Jones). The results showed sizable disparate treatment effects, with the White-named applicants receiving fifty percent more callback interviews.[150]

In both studies, it is possible that explicit bias drove some of the disparate treatment. But implicit bias is also implicated because decision

---

147. Sturm reports substantial improvements in gender disparities following these changes. *See id.* at 498 ("By 1995, 23% of senior managers were women, the percentage of women admitted to partner rose from 8% in 1991 to 21% . . . .").

148. *See, e.g.*, Jerry Kang, *Cyber-race*, 113 HARV. L. REV. 1131, 1133-34 (2000) (providing the anecdote of a minority employing a buying agent to purchase a car as an example of cloaking because it "remove[s] racialized negotiations from the car buying ritual"); *see also* Kang, *supra* note 12, at 1499-1504 (describing racial mapping as component of racial mechanics model).

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

150. *See* Marianne Bertrand & Sendhil Mullainathan, *Are Emily And Greg More Employable Than Lakisha And Jamal? A Field Experiment On Labor Market Discrimination* 2 (Nat'l Bureau of Econ. Research, Working Paper No. 9873, 2003). For a detailed description of the methodology and results in the law reviews, in the context of a behavioral realist project, see Kang, *supra* note 12, at 1515-17. For a similar study on gender, see David Neumark, Roy J. Bank & Kyle D. Van Nort, *Sex Discrimination in Restaurant Hiring: An Audit Study* 710 (Nat'l Bureau of Econ. Research, Working Paper No. 5024, 1995) (finding gender discrimination in interview callbacks at high priced restaurants). *See generally* P.A. Riach & J. Rich, *Field Experiments of Discrimination in the Market Place*, 112 ECON. J. 480 (2002) (summarizing field experiments over three decades and ten different nations).

makers should be expected to prefer the best musician or employee irrespective of gender and race. Preventing irrelevant information from influencing decision making is a simple and smart way to measure merit more accurately.

Social category cloaking can be implemented in both educational and employment settings. For example, schools can conduct blind grading of student work.[151] In the hiring context, we can remove names, pictures, and other category-signaling data by temporarily assigning candidates pseudonyms. In the marketplace, given evidence of discrimination in car[152] and real estate purchases,[153] pseudonymous credit credentialing, purchase intermediaries, and auction systems could make big-ticket sales possible while cloaking identity.[154] Cyberspace and virtual worlds offer still other avenues for interaction while cloaking identity.[155]

Our call for race and gender cloaking is not naïve. Creating successful cloaking regimes is difficult given the myriad ways that social category membership is signaled.[156] It also does nothing to correct for accumulated discrimination.[157] Finally, cloaking will not magically erase group differences in performance: one sees, for instance, gender and race differences in law school examinations even when they are graded blind. None of this,

---

151. In this discussion, we do not take any position for or against time-pressured, closed-book, in-class examinations as compared to, for example, open-book, day-long, take home exams. Different exam formats may produce different disparities; for example, the former exam format may produce a greater gender gap than the latter. Insisting that the former exam format is the better measure of merit invites a critique from the "merit as fraud" position: What is your evidence?

152. *See* Fiona Scott Morton, Florian Zettelmeyer & Jorge Silva-Risso, *Consumer Information and Discrimination: Does the Internet Affect the Pricing of New Cars to Women and Minorities?*, 1 QUANTITATIVE MKTG. & ECON. 65, 68, 91 (2003) (finding that Internet purchasing of cars eliminates most of the race premium (all less than 2.3%, depending on which variables are controlled) that minorities otherwise end up paying).

153. *See, e.g.*, DOUGLAS S. MASSEY & NANCY A. DENTON, AMERICAN APARTHEID: SEGREGATION AND THE MAKING OF THE UNDERCLASS 99-100 (1993) (reporting studies by HUD and George Galster showing racial discrimination in sales and rental markets).

154. *See* James Bandler, *Harvard Ponders A River Crossing; Some Graduate Facilities May Relocate To Allston*, BOSTON GLOBE, Aug. 22, 1999, at B1 (describing Harvard's purchase through a private broker so that sellers would not raise prices).

155. For an extensive discussion of how cyberspace can be used to "abolish" race in market transactions, see Kang, *supra* note 148, at 1133-35, 1154-60 (including discussion of pseudonymous credentialing systems).

156. *See, e.g.*, Kang, *supra* note 155, at 1156-60.

157. "In principle, blinding appears to qualify as a fool-proof method of avoiding unintended discrimination" but in practice, blinding can only account for "stigmatizing attribute[s]" and fails to account for other manifestations of discrimination. Anthony G. Greenwald & Mahzarin R. Banaji, *Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes*, 102 PSYCHOL. REV. 4, 19 (1995). For example, orchestras now have candidates audition out of the sight of evaluators to remove gender biasing cues. However, to the extent male performers have accrued the benefits of past discrimination resulting in a distinguishable difference in ability (e.g., opportunity to attend Juilliard), men will continue to "maintain relative success" because the "[d]isadvantages [women have] inherited from past discrimination are not undone by blinding." *Id.*

however, refutes our claim that cloaking social category produces more accurate evaluations of merit.

Our recommendation is distinct from the standard demand for color or gender-blindness. As typically formulated, this is a moral exhortation for actors in some delimited public sphere to simply ignore race and gender when making decisions. In other words, even though the social category information is perceived, actors are encouraged to apply cognitive effort to ignore it. As we have pointed out, an explicit commitment to be cognitively blind hardly guarantees neutral behavior. We therefore recommend the proactive strategy of self-awareness (to learn that we are not color or gender blind) and the prophylactic strategy of removing the information from even entering the cognitive decision making process. Once such information seeps in, mere exhortations to ignore it must be viewed skeptically. We must be behaviorally realistic.

To avoid misunderstanding, we underscore that our analysis assumes that the social category is in fact irrelevant to merit. If that is not the case—for example, gender would be relevant to picking an undercover agent to infiltrate a gang of female bikers—then cloaking social category would be irrational. We also note that a more accurate measurement of merit does not have to be the sole driver of a selection system. Other considerations, such as corrective and distributive justice, which can extend beyond the scope of "fair measures," can warrant adjustments on the basis of race and gender. In such cases, we recommend a two-step selection process, in which identity is cloaked in the first-step to gauge "merit," but then the veil is lifted in a second-step to evaluate any additional considerations. Our call for cloaking to prevent implicit-bias induced discrimination is agnostic about whether these additional considerations are warranted.

*b. Discount the Emphasis on Traditional Interview*

Another way to reduce perceiver-side bias is to decrease subjective discretion in the merit measurement. Interviews are extraordinarily subjective, and for the past four decades, evidence has mounted that making decisions based on interviews produces worse outcomes than arriving at them via the paper record.[158] On the basis of such evidence, the Princeton philosophy department decided decades ago not to interview for tenure-track jobs. It makes its judgments solely on the paper record.

Recently, we have also learned that interview interactions can be influenced by implicit bias. For example, Allen McConnell and Jill Leibold

158. *See* Richard D. Arvey, *Unfair Discrimination in the Employment Interview: Legal and Psychological Aspects*, 86 PSYCHOL. BULL. 736, 759-60 (1979); Linda L. Frank & J. Richard Hackman, *Effects of Interviewer-Interviewee Similarity on Interviewer Objectivity in College Admissions Interviews*, 60 J. APPLIED PSYCHOL. 356 (1975); Eugene C. Mayfield, *The Selection Interview: A Re-evaluation of Published Research*, 17 PERSONNEL PSYCHOL. 239 (1964); Lynn Ulrich & Don Trumbo, *The Selection Interview Since 1949*, 63 PSYCHOL. BULL. 100 (1965).

demonstrated the linkage between IAT results and intergroup behavior. In this experiment, White participants were required to interact with White and Black confederates under scripted conditions. Trained third parties blind to the purpose of the experiment and participants' bias scores coded the participants' body language during the interactions to measure overall friendliness. The trained observers scored items such as eye contact, forward body lean, arm positioning, and number of speech errors.[159] The higher the implicit bias, the more awkward was the social interaction.[160]

Although awkwardness might seem trivial, prior research confirms that awkwardness leads to worse interviews. Early research by Carl Word, Mark Zanna, and Joel Cooper demonstrated that when White interviewers (confederates) were trained to perform unfriendly nonverbal behavior — the sort that has now been correlated with higher implicit bias against racial minorities[161]—in front of *White* interviewees (study participants), those interviewees gave worse interviews, as measured objectively by third parties blind to the purpose of the experiment.[162] A positive feedback loop typically creates a vicious cycle in which the unfriendly behavior is replicated by the target, and the social interaction degrades.[163] The consequences are weighty. For instance, in law firm hiring, interviews are mostly a check of "personality" or "fit."[164] Accordingly, an uncomfortable interview can make all the difference.

Obviously, in various hiring contexts—for example, at promotion as opposed to initial hiring[165]—it will be infeasible to remove interviews or evaluations based on social interactions entirely. But its weighting relative to other performance measures can be decreased. In addition, by interviewing an extensive pool of potential candidates and evaluating them in

159. Other observed behaviors included abruptness, general comfort level, degree of laughter, forward body lean, direction of body facing experimenter, openness of arms, expressiveness of arms, distance between seats, speaking time, number of smiles, number of speech hesitations, number of fidgeting body movements, and number of extemporaneous social comments. 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 (2001).

160. *See id.* at 439.

161. *See* Carl O. Word et al., *The Nonverbal Mediation of Self-Fulfilling Prophecies in Interracial Interaction*, 10 J. EXPERIMENTAL SOC. PSYCHOL. 109 (1974).

162. *See id.*

163. *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).

164. *See* Wilkins & Gulati, *supra* note 140, at 547-48, 557-59.

165. *Cf.* Robert E. Thomas & Bruce Louis Rich, *Under the Radar: The Resistance of Promotion Biases to Market Economic Forces*, 55 SYRACUSE L. REV. 301, 303-05 (2005) (distinguishing entry-level labor market and promotion labor market in a hierarchical organization, and explaining that merit measurements in the latter market are much more difficult to evaluate).

accordance with well-specified, pre-set guidelines, decision makers can diminish interview subjectivity.[166]

Recent research has confirmed the value of adopting such approaches. Using subsequent ratings of job performance or training performance as the criteria for measuring the validity of interviews, studies have shown that behavior-based, structured interviews do better than unstructured interviews at predicting on-the-job success. In other words, the more unstructured the interview, and hence the greater the chance for individual preferences to play a role in decision-making, the poorer the outcome. The paper record and structured interviews combat these biases better than standard interviews alone.[167]

*c. Remove Stereotype-Threat Triggers*

Up to now, we have discussed how society might intervene on the perceiver side to produce fairer measures of largely subjective evaluations. What might be done on the target side, regarding stereotype threat, which warps even objective evaluations? We know of no sure-fire solutions to this problem, partly because researchers have not yet identified the precise cognitive mechanisms of the threat.[168] On the limited information we have, we make a few modest suggestions.

First, stereotype threat can sometimes be decreased by telling test-takers that the stereotype is irrelevant. One can simply assert that the test is not "diagnostic" of the stereotyped trait. For example, one can explain that the test is an "exercise" and not a test of native "intelligence." Under these conditions, the Black undergraduates in the Steele and Aronson

166. One example comes from a class action suit by female employees against Home Depot, which ended in a consent decree. The settlement required structural changes that allowed employees to express their job preferences in a computer database, which generated automatic lists of qualified employees for any available promotion. It also required that decision makers interview at least three candidates for each position pursuant to a guided script. By various accounts, these reforms were successful. *See generally* Tristin K. Green, *Targeting Workplace Context: Title VII As A Tool for Institutional Reform*, 72 FORDHAM L. REV. 659, 684-85 (2003) (noting that the consent decree was lifted a year earlier than planned). *But see* Michael Selmi, *The Price of Discrimination: The Nature of Class Action Employment Discrimination Litigation and Its Effects*, 81 TEXAS L. REV. 1249, 1285-88 (2003) (expressing skepticism).

167. Elaine D. Pulakos & Neal Schmitt, *Experience-Based and Situational Interview Questions: Studies of Validity*, 48 PERSONNEL PSYCHOL. 289, 306 (1995) ("Interviews in which applicants are asked the same job-relevant questions and whose answers are evaluated using specifically anchored rating scales are likely to produce higher levels of validity than other types of interviews."). *Cf.* Uhlmann & Cohen, *supra* note 141, at 479 (discussing evidence that a prior commitment to specific merit standards erased gender discrimination caused by the tendency to redefine merit "to fit the idiosyncratic qualifications of applicants who belonged to favored groups").

168. *See supra* note 125. There have been some promising advances, however, on this front. *See, e.g.*, Jean-Claude Croizet et al., *Stereotype Threat Undermines Intellectual Performance by Triggering a Disruptive Mental Load*, 30 PERSONALITY AND SOC. PSYCHOL. BULL. 721, 726 (2004).

experiments showed no depression of test scores.[169] Also, one can proclaim that regardless of the general diagnostic or nondiagnostic nature of a particular type of exam, this *particular* test shows no differences among social categories. For example, Steven Spencer and his colleagues demonstrated that women were generally subject to stereotype threat on math tests. However, when they were instructed that this particular test showed no gender difference, the stereotype threat disappeared.[170] Any such proclamations would have to be credible. If it is well established and well known that the LSAT features a robust racial disparity, a proctor's tepid denial of that fact before the examination may not neutralize the stereotype threat.[171]

Second, a person's particular theory of intelligence might alter her vulnerability to stereotype threat. If one believes that intelligence is mostly natural, given, and fixed, one is more susceptible. By contrast, if one believes that intelligence is mostly nurtured, achieved, and malleable, one is more resistant to stereotype threat.[172] Recently, Catherine Good and her colleagues confirmed these laboratory findings in a real-world field experiment in the state of Texas. By giving seventh grade students mentors who emphasized an intelligence-is-malleable message instead of a control message about drug abuse, these researchers prevented stereotype-threat from depressing the performance of seventh grade girls on standardized math exams.[173] The effect sizes were substantial. On these standardized tests, a score below 70 is similar to failing. In the control condition, the girls scored a mean of 74 compared to the boys' mean score of 81.55. By contrast, in the treatment condition (which included two different kinds of messages about intelligence malleability), the girls averaged 84.06, with the boys actually scoring slightly less.[174] This slight difference was not statistically significant. In other words, under the intelligence-is-malleable condition, the girls scored on average the same as the boys and showed no stereotype-threat depression in performance.

Third, there may be ways to disarm environmental triggers of the stereotype threat. Triggers can range from extremely subtle (e.g., being a numerical minority in the examination room) to obvious (e.g., being

169. *See* Claude M. Steele, *A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance*, 52 AM. PSYCHOLOGIST. 613 (1997).

170. *See* Steven J. Spencer et al., *Stereotype Threat and Women's Math Performance*, 35 J. EXPERIMENTAL SOC. PSYCHOL. 4, 11, 13, 17 (1999).

171. Difficult ethical questions are raised by this potential strategy. For instance, under what conditions may a professor state that his exams do not show any race disparities in order to minimize stereotype threat? Can she say so without actually checking, knowing that the simple statement might create a self-fulfilling prophecy?

172. *See* Joshua Aronson et al., *Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence*, 38 J. EXPERIMENTAL SOC. PSYCHOL. 113 (2002).

173. *See* Catherine Good et al., *Improving Adolescents' Standardized Test Performance: An Intervention to Reduce the Effects of Stereotype Threat*, 24 APPLIED DEV. PSYCHOL. 645, 655-56 (2003).

174. *See id.* at 656.

explicitly told of stereotypes regarding performance).[175] In certain cases, such as in high stakes testing, preventing the trigger may be impossible.[176] But in other cases, one might change or avoid certain media and environmental cues to decrease the likelihood of stereotype activation. For example, in one study by Paul Davies and his colleagues about television commercials, male and female participants were exposed to a set of stereotypic or counterstereotypic television commercials.[177] In addition to a set of commercials common to both groups, there was a stereotypic set that featured a commercial with a woman excited about acne medicine, and another excited about brownies. A counterstereotypic set featured one commercial with a woman demonstrating automotive expertise, and another demonstrating healthcare expertise. After viewing these commercials, the participants took a difficult math test. When primed with counterstereotypic ads, women and men scored equally well ($M = 31\%$ for women; $M = 34\%$ for men). Yet in the stereotypic condition, women performed far worse than the men ($M = 19\%$ and $M = 39\%$ respectively).[178] Such studies identify the issue—the importance of environmental cues—but do not provide concrete, foolproof methods to avoid stereotype activation. We simply point out that as our scientific understanding improves, a wide range of strategies may become relevant, including media interventions[179] and nonconscious priming.[180]

### 3. *Correct Mismeasures: Breaking Ties in Favor of Bias Targets*

Suppose two candidates end up in a tie for some position. However, one candidate belongs to a social category that suffers from implicit bias, which may have depressed her merit scores. In such a case, instead of flipping a coin, the tie should be broken in favor of that candidate. Our justification is not corrective justice or moral desert. Rather, the justification is

---

175. *See* Smith, *supra* note 125, at 181.

176. Some laboratory studies found that the race/ethnicity or gender of the proctor could influence whether the stereotype threat was activated. *See, e.g.*, D.M. Marx & J.S. Roman, *Female Role Models: Protecting Women's Math Test Performance*, 28 PERSONALITY AND SOC. PSYCHOL. BULL. 1183 (2002) (gender); A.M. Walters, J.A. Shepperd & L.M. Brown, The Effect of Test Administrator Ethnicity on Test Performance, (2003) (unpublished manuscript). The Educational Testing Service performed a real-world study, and although it had various methodological limitations, it could not replicate these findings in operational settings. *See* WALTERS, ET AL., *supra* note 122, at 34.

177. Paul G. Davies et al., *Consuming Images: How Television Commercials That Elicit Stereotype Threat Can Restrain Women Academically and Professionally*, 28 PERSONALITY & SOC. PSYCHOL. BULL. 1615, 1619 (2002).

178. *Id.* at 1620 ($p < 0.01$).

179. *See generally* Kang, *supra* note 12, at 1549-63, 1579-85 (discussing the harm of local news and the possibility of debiasing public service announcements as a disinfection strategy).

180. *See, e.g.*, Kai Sassenberg & Gordon B. Moskowitz, *Don't Stereotype, Think Different! Overcoming Automatic Stereotype Activation by Mindset Priming*, 41 J. EXPERIMENTAL SOC. PSYCHOL. 506 (2005) (providing some evidence that unconscious priming to think creatively decreases activation of stereotypes).

that any candidate who registers a tie on an instrument that is biased against her is likely to be the stronger candidate. One cannot be certain in any specific case, but on average, this approach will measure merit more accurately.

Operationalizing this simple insight is, unfortunately, complex. First, many implicit biases exist, not only those based on immutable social categories. For example, there is ingroup bias based on school attended, geographic region, physical similarity, or shared culture (jokes, knowledge, artistic interests etc.). To obtain the most accurate merit measure, one should account for all these biases in any tie-breaking algorithm. But the attempt to be comprehensive increases both the cost of the selection process and the likelihood of cross-cutting biases on the same or both sides of the equation. For example, when a younger East coast woman with average looks is tied with an older, short, but attractive, Midwestern man for a Silicon Valley job, how should the tie be broken? In practice, then, we should focus on only the most consequential implicit biases, which future research can help identify more precisely. In addition, other goals besides the most accurate measure of merit, such as corrective justice, distributive justice, or antisubordination based on immutable traits, can inform which biases are most important to counter.

Second, it may be difficult to determine what constitutes a "tie." If merit is measured qualitatively, ties occur when the evaluator feels that it is a genuine toss-up between the two candidates in terms of merit. If merit is measured quantitatively, ties occur when scores are identical. But notice that even identical scores typically represent some banding. For example, even if two candidates receive an identical LSAT score of 155, they may not have answered the same number of questions correctly. If the LSAT score were computed to another significant digit, one candidate may have received 155.3, whereas the other 155.4. Nevertheless, the scores are both reported and weighted as 155, a numerical tie. Such aggregation raises the question of whether a score of 155 should be treated as a tie with a score of 154, and so on. Recall the facts of *Johnson v. Transportation Agency*,[181] in which Paul Johnson sued on the basis of his dispatcher examination score of 75 being two points higher than that of Diane Joyce's score of 73.[182] Should these results have been considered a tie?

Answering this question intelligently requires a lengthy and difficult analysis, which we do not attempt here. We simply observe that in order to account for random measurement error, one should consider the test's validity, reliability, standard error of measurement, standard error of

181. 480 U.S. 616 (1987)
182. *See id.* at 624.

difference,[183] confidence level desired that two different observed scores reflect differences in true scores, and the costs of Type I and Type II errors.[184] Also, to account for systematic error induced by stereotype threat, boost, and lift, one should consider mean effect sizes of these testing phenomena,[185] their variance, and the likelihood the stereotype phenomenon was triggered. An answer cannot be determined simply by applying uncontroversial statistical techniques. Rather, choices that reflect values expressed in uncertainty must be made. For purposes of this paper, we stake out the most conservative position of counting only identically reported scores as a tie.

The fair measures we have previously recommended, such as avoiding unstructured interviews, were not facially race- or gender-based. Accordingly, they are not especially vulnerable to legal challenges under the Equal Protection Clause or Title VII.[186] However, the same cannot be said of our tie-breaker recommendation, even in its most conservative formulation, because it is explicitly race- and gender-conscious. That said, we believe that tie-breakers established on the grounds that they are the more accurate measures of merit can withstand legal scrutiny. The strongest legal authority comes from the line of cases that have upheld tie-breaking in favor of racial minorities under constitutional and Title VII challenge[187] in the context of civil service exams for firefighters and law enforcers. The governmental objective that was accepted in these cases was remedying past discrimination. Our argument for breaking ties is stronger because the objective is not to roughly remedy past discrimination, but to prevent present acts of inaccurate measurement. Also, if some form of banding is adopted in the definition of a "tie," that could not be said to constitute

183. For a discussion of these standard error measurements in the law reviews, see Selmi, *supra* note 81, at 1272.

184. *See* Sheldon Zedeck et. al., *Sliding Bands: An Alternative to Top-down Selection*, *in* FAIR EMPLOYMENT STRATEGIES IN HUMAN RESOURCE MANAGEMENT 222, 230-32 (Richard S. Barrett ed., 1996) (providing a sliding band methodology, with width of the band set as a function of necessary confidence level and risk analysis). Type I errors suppose that there is a difference between test scores when in fact there is not. Type II errors suppose that there is no difference when in fact there is. *See id.* at 232. For airing of some controversies around banding, see Michael A. Campion et al., *The Controversy Over Score Banding in Personnel Selection: Answers to 10 Key Questions*, 54 PERSONNEL PSYCHOL. 149 (2001); Wayne F. Cascio et. al., *Twenty Issues and Answers About Sliding Bands*, 8 HUMAN PERF. 227, 238 (1995) (rejecting, among other things, that banding is race-norming).

185. *See* Walton & Cohen, *supra* note 133, at 463 (comparing studies that affirmatively refute link between test and stereotypes to those that do not; $d = 0.24$; $p < 0.0001$).

186. *Cf.* Hayden v. County of Nassau, 180 F.3d 42, 53 (2d Cir. 1999) (holding that designing tests to have less disparate impact does not violate equal protection or Title VII).

187. *See, e.g.*, Cotter v. City of Boston, 323 F.3d 160 (1st Cir. 2003); Chicago Firefighters Local 2 v. City of Chicago, 249 F.3d 649, 657-58 (7th Cir. 2001); Boston Police Superior Officers Fed'n v. City of Boston, 147 F.3d 13 (1st Cir. 1998); Officers for Justice v. Civil Serv. Comm'n of City and County of San Francisco, 979 F.2d 721, 728 (9th Cir. 1992).

race-norming,[188] which is expressly proscribed by section 106 of the Civil Rights Act of 1991.[189] Instead, as Judge Richard Posner put it, banding is an "unquestioned method of simplifying scoring by eliminating meaningless gradations."[190]

## III
## DECREASING BIAS

### *A. Conventional Thinking: Social Contact Hypothesis*

An asserted benefit of affirmative action has been integration within the workplace and academic institutions that undermines stereotypes and prejudice against disadvantaged groups. To be sure, some proponents care only about affirmative action's capacity to redistribute material resources from privileged to subordinated groups without either hope or regard for debiasing attitudes and beliefs. But, most affirmative action defenders have trumpeted not only material but also psychological benefits. Indeed, material redistribution without decreasing bias might be only a short-term solution because bias produces discrimination, which reproduces material inequalities.

A principal mechanism for psychological change is the "social contact hypothesis." This theory suggests that when individuals of different social categories interact face-to-face under certain conditions, their stereotypes and prejudice will be tempered. Since the 1950s when the social contact hypothesis was first proposed, social psychologists have distilled the conditions that contribute to a debiasing environment. People must be: (1) exposed to disconfirming data; (2) interact with others of equal status; (3) cooperate; (4) engage in non-superficial contact; and (5) receive clear norms in favor of equality.[191]

Many opponents of affirmative action vigorously dispute that affirmative action improves attitudes in practice. For example, champions of colorblindness argue that by being conscious of race, affirmative action strengthens identification with and resentment across race. This is a balkanization story frequently told by conservative commentators and

188. *See, e.g.*, *Chicago Firefighters Local 2*, 249 F.3d at 656 (Posner, J.) (deciding question of first impression).

189. *See* 42 U.S.C.A. § 2000e-2(l) ("Prohibition of discriminatory use of test scores. It shall be an unlawful employment practice for a respondent, in connection with the selection or referral of applicants or candidates for employment or promotion, to adjust the scores of, use different cutoff scores for, or otherwise alter the results of, employment related tests on the basis of race, color, religion, sex, or national origin.").

190. *Chicago Firefighters Local 2*, 249 F.3d at 656.

191. *See* Norman Miller & Marilynn B. Brewer, *The Social Psychology of Desegregation: An Introduction, in* GROUPS IN CONTACT: THE PSYCHOLOGY OF DESEGREGATION 1, 2 (Norman Miller & Marilynn B. Brewer eds., 1984).

skeptical judges.[192] Further, many commentators, even those sympathetic to affirmative action's goals, point out the difficulty in achieving the five conditions listed above.[193] For example, by flexing traditional merit standards too much, we might create unequal engagements, which perversely bolster, not undermine, stereotypes of inferiority.[194] The social contact hypothesis has also been challenged by those farther out on the Left as both ineffective and assimilationist.[195] What does implicit social cognition add to the debate?

### *B. Behavioral Realism: Revising Mechanisms for Decreasing Bias*

#### *1. Social Contact Hypothesis (SCH)*

Thomas Pettigrew and Linda Tropp have recently provided a definitive meta-analytic test of the SCH, which demonstrates that intergroup

---

192. In his *Adarand* concurrence, Justice Thomas warned that the conscious use of race "stamp[s] minorities with a badge of inferiority and may cause them to develop dependencies or to adopt an attitude that they are 'entitled' to preferences." Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 241 (1995). At the same time, affirmative action programs "engender attitudes of superiority [in Whites] or, alternatively, provoke resentment [because of the belief] they have been wronged by the government's use of race." *Id.* at 241.

193. *See* Krieger, *Civil Rights Perestroika, supra* note 80, at 1263-70.

194. *See id.* at 1263 (suggesting consensus view that preferential affirmative action can trigger resentment and stereotypic evaluations). *Cf.* Richard H. Sander, *A Systemic Analysis of Affirmative Action in American Law Schools*, 57 STAN. L. REV. 367, 371-72 (2004) (arguing affirmative action has created a system where "black law applicants end up at schools where they will struggle academically and fail at higher rates than they would in the absence of preferences" resulting in blacks' lower bar passage rates and depressed job opportunities upon graduation). Professor Sander's article prompted numerous responses. *See* Ian Ayres & Richard Brooks, *Does Affirmative Action Reduce the Number of Black Lawyers?*, 57 STAN. L. REV. 1807, 1809 (2005) (finding that "the elimination of affirmative action would reduce the number of [black] lawyers"); David L. Chambers et. al., *The Real Impact of Eliminating Affirmative Action in American Law Schools: An Empirical Critique of Richard Sander's Study*, 57 STAN. L. REV. 1855, 1857 (2005) (asserting "[t]he conclusions in [Sander's article] rest on a series of statistical errors, oversights, and implausible assumptions"); Kevin R. Johnson & Angela Onwuachi-Willig, *Cry Me a River: The Limits of "A Systemic Analysis of Affirmative Action in American Law Schools,"* 7 AFR.-AM. L. & POL'Y REP. 1, 4 (2005) (arguing that Sander "neglects to account for the well-documented hostile environment . . . in law school and how it may adversely affect academic performance"); David B. Wilkins, *A Systemic Response to Systemic Disadvantage: A Response to Sander*, 57 STAN. L. REV. 1915, 1918 (2005) (arguing Sander fails "to prove that grades are more important than law school prestige for those black law students who actually become lawyers").

195. Richard Delgado, for example, rejects the usefulness of social contact and instead encourages direct confrontation as the best strategy to decrease bias. *See, e.g.*, Richard Delgado, *Rodrigo's Twelfth Chronicle: The Problem of the Shanty*, 85 GEO. L.J. 667, 682 (1997); Richard Delgado, Book Review, *Stark Karst*, 93 MICH. L. REV. 1460, 1470-71 (1995). *Cf.* Stephen M. Feldman, *Whose Common Good? Racism in the Political Community*, 80 GEO. L.J. 1835, 1859-60 (1992) (suggesting that social contact theory is self-contradictory since creating equal status interactions is difficult in a racist society). According to one recent study, confrontation may create guilt, especially among those with low explicit prejudice, but not attitude change. *See* Alexander M. Czopp & Margo J. Monteith, *Confronting Prejudice (Literally): Reactions to Confrontations of Racial and Gender Bias*, 29 PERSONALITY & SOC. PSYCHOL. BULL. 532 (2003).

interaction decreases prejudice.[196] Reviewing 515 studies using 713 independent samples that encompassed a quarter million people from 38 nations, they found that intergroup contact correlates negatively with prejudice (average $r$ =-0.215; $p < .0001$).[197] The more careful and rigorous the study, the larger the effect seen.[198] Also, none of the five elements outlined in the previous section were strictly necessary. Finally, the effect of a broad range of intergroup contact situations showed generalization of the improved attitude to the target's entire outgroup.

The researchers specifically examined whether the causal sequence might be operating in reverse, i.e., whether less prejudiced people sought out greater intergroup contact. By distinguishing studies by the degree of choice persons had in engaging in such contact, the researchers tested whether selection bias was a significant problem. Holding relevant variables constant, they saw no significant correlation between "choice" to interact and the effect size in prejudice decrease ($r$ = .005, $p$ = .89).[199] The broad lesson, then, is that social integration works.

The more recent science of implicit social cognition (ISC) helps identify more precisely the conditions in which bias is most likely to be reduced. Various ISC studies provide interesting results on this front. For example, greater intergroup contact with members of an outgroup has been found to be associated with lower IAT attitudinal bias. In one study, Christopher Aberson and his colleagues asked White participants to take the race IAT and report the number of their close outgroup friends: African-Americans in one experiment and Latinos in another.[200] On the basis of their answers, the participants were put into one of two categories: "no friends" or "friends." The researchers found negative correlations between the number of interracial friendships and level of implicit bias.[201] Reflecting the general pattern of dissociation, the friendship measure had no correlation to measures of explicit bias.[202]

---

196. Thomas F. Pettigrew & Linda R. Tropp, *A Meta-Analytic Test of Intergroup Contact Theory*, 90 J. PERSONALITY & SOC. PSYCHOL. 751 (2006).

197. In this careful paper, the researchers also examined various threats to validity including the causal sequence problem (due to potential selection bias of individuals in the experiments), the file drawer problem, and the generalization of effects problem. They concluded that these results were not artifacts of participant selection or publication bias; further, the effects typically generalized beyond the specific participants in the contact situation. "Not only do attitudes toward the immediate participants usually become more favorable, but so do attitudes toward the entire outgroup, outgroup members in other situations, and even outgroups *not* involved in the contact." *Id.*

198. Of the seventy-seven samples that used the most rigorous measures of contact and prejudice, as well as adequate controls, the mean effect was far stronger ($r$ = -.323; $p < .0001$). *Id.*

199. *See id.*

200. Christopher L. Aberson, Carl Shoemaker & Christina Tomolillo, *Implicit Bias and Contact: The Role of Interethnic Friendships*, 144 J. SOC. PSYCHOL. 335 (2004).

201. *See id.* at 340, 343 (African Americans and Latinos, respectively).

202. *See id.* at 341, 344 (African Americans and Latinos, respectively). One weakness of this study is that it used self-reports of "close friends" without guidance or definition of the term.

In another study, Andreas Olsson and colleagues explored how fear is learned and extinguished by measuring skin conductance on both Black and White participants.[203] They demonstrated that fear responses to both ingroup and outgroup member pictures could be learned (by receiving small electric shocks). These fear responses could also be extinguished, by showing the pictures again but without the shocks; however, participants retained "learned" fear for a longer time period when that fear was associated with a racial other.[204] Researchers found, however, one mediator: this response bias (of retaining fear of racial others longer) was negatively correlated with the number of outgroup romantic partners that the participant had ($r$ =-0.29, $p < 0.05$).[205]

While the studies just cited lend general support to the SCH, other ISC research complicates the story. For example, in 2002, Joshua Correll confirmed earlier findings of "shooter bias."[206] He created a video game that flashed photographs of a White or Black individual holding either a gun or harmless object (such as a cell phone).[207] Participants were told to decide to shoot if they saw a gun or to refrain from shooting if they did not see a gun. Under severe time pressure, participants made errors, but they were not randomly distributed. Instead, participants more often mistook a Black target as armed when he was unarmed (false alarms);[208] conversely, they more often mistook a White target as unarmed when he was armed (misses).[209] Such shooter bias was found in both Black and White participants.[210] Still more perplexing—and connected to the SCH—is that shooter bias was correlated with the amount of contact that the participants claimed to have had with African Americans. If we believe these self-reports, we have at least one study that shows that increased interracial contact produces a greater tendency to "shoot" African Americans.

One way to reconcile such contrary findings is to distinguish an "attitude," which is a feeling or preference that carries with it an affective overtone, from a "stereotype," which reflects a belief about a group.[211] Having a

---

203. *See* Olsson, *supra* note 65, at 785.

204. *See id.* at 786 (suggesting that this could be one reason why we have "more negative evaluations of the outgroup").

205. *See id.* ("Specifically, the conditioning bias to outgroup faces was negatively correlated with the reported number of outgroup, relative to ingroup, romantic partners.").

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

207. 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 experimental setup).

208. *See id.* at 1319. This finding was statistically significant at $p < 0.02$, with outlier images that produced too many errors thrown out.

209. *See id.* This finding was significant at $p < 0.001$.

210. *See id.* at 1325.

211. An example of an attitude is feeling positively toward the category of flowers. An example of a stereotype is the belief that the category of rattling snakes may be poisonous.

positive attitude toward a category does not preclude holding a negative stereotype about that category. For example, most people exhibit an in-group favoritism in their implicit attitudes. That is less true, however, with implicit stereotypes. For example, women show an implicit attitudinal preference for females over males,[212] but they nonetheless show an implicit stereotype linking females closer to family than career.[213] It may be that the "facts" as we perceive them in our daily lives affect us in ways that are hard to set aside—even though being a member of the group encourages us to reject the stereotype. Interestingly, subsequent shooter studies have demonstrated that stereotypes, not attitudes, drive this bias.[214] Thus, increased contact with African Americans may make implicit attitudes more favorable, but any such improvement would not impact shooter bias, which is stereotype driven.

In sum, research supports the value of intergroup contact to ameliorate negative attitudes (also called "prejudice"). However, intergroup contact may not counteract negative stereotypes. Clarification awaits future research.

### 2. *Countertypical Exemplars*

In addition to refining the social contact hypothesis, ISC suggests a slightly different mechanism for decreasing bias—exposure to "countertypical," by which we mean counterattitudinal or counterstereotypic, exemplars. Consider the following three studies.

*Countertypical celebrities.* Nilanjana Dasgupta and Anthony Greenwald found that implicit attitudes could be changed simply by

---

212. *See, e.g.*, 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, 506 (2004).

213. Brian A. Nosek, Mahzarin R. Banaji & Anthony G. Greenwald, *Harvesting Implicit Group Attitudes and Beliefs from a Demonstration Web Site*, 6 GROUP DYNAMICS 101, 108-09 (2002)

214. *See* Charles M. Judd et al., *Automatic Stereotypes vs. Automatic Prejudice: Sorting Out the Possibilities in the Payne (2001) Weapon Paradigm*, 40 J. EXPERIMENTAL SOC. PSYCHOL. 75, 80 (2004). Treatments that decrease shooter bias have been consistent with this diagnosis. Ashby Plant and colleagues demonstrated that shooter bias was eliminated for participants who were repeatedly exposed to Black and White faces where there was no relation to race and the presence of a gun. E. Ashby Plant et. al., *Eliminating Automatic Racial Bias: Making Race Non-Diagnostic for Responses to Criminal Suspects*, 41 J. EXPERIMENTAL SOC. PSYCHOL. 141, 147 (2005). The elimination of bias persisted after a twenty-four hour period and also led to an "inhibition of racial concepts" (depressed ability to complete race-related words). *Id.* at 149, 152-53. In contrast, shooter bias was not eliminated for participants who were repeatedly exposed to Black and White faces when the Black face was associated with the presence of a gun at a 70% rate. *Id.* at 150-51; *see also* E. Ashby Plant & B. Michelle Peruche, *The Consequences of Race for Police Officers' Responses to Criminal Suspects*, 16 PSYCHOL. SCI. 180 (2005) (demonstrating that police officers exhibited shooter bias but that it was eliminated after multiple exposures to Black and White faces when there was no relation to race and the presence of a gun).

exposing people to pictures of particular individuals.[215] First, researchers gave participants a "general knowledge" questionnaire. For the pro-Black condition group, the questionnaire contained names and images of positive Black exemplars, such as Martin Luther King, Jr. and Denzel Washington, and negative White exemplars, such as Jeffrey Dahmer and Howard Stern.[216] For the pro-White condition group, the questionnaire contained names and images of negative Black exemplars, such as Louis Farrakhan and Mike Tyson, and names and images of positive White exemplars, such as John F. Kennedy and Peter Jennings.[217] After finishing the questionnaire, participants took an IAT and then completed a survey of racial bias. Although the type of questionnaire had no impact on participants' explicit bias as measured by self-reports, it had a surprisingly significant effect on IAT scores. Participants in the pro-Black condition reduced their implicit bias by more than half.[218] The reduction persisted for a full day as measured by a follow-up test.[219]

*Countertypical visualizations.* Do we actually need to look at photographs of countertypical historical figures to reduce our implicit bias? Perhaps we can reduce implicit bias simply by imagining the right exemplar. After all, Isabelle Klein and colleagues have shown that several brain regions in the striate cortex show exactly the same pattern of activation regardless of whether the individual is actually seeing or merely imagining a specific object.[220] In fact, Irene Blair and colleagues demonstrated that mental imagery, through a form of self-priming, moderated implicit stereotypes.[221] A group of participants was instructed to spend a few minutes

---

215. 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. PSYCH. 800, 807 (2001).

216. The complete list of positive Black images included Martin Luther King, Jr., Jesse Jackson, Colin Powell, Denzel Washington, Eddie Murphy, Michael Jordan, Tiger Woods, Will Smith, Bill Cosby, and Gregory Hines. For negative White images, Dasgupta and Greenwald used Ted Bundy, Jeffrey Dahmer, Timothy McVeigh, Charles Manson, Al Capone, Ted Kaczynski, Terry Nichols, Howard Stern, John Gotti, and John Dillinger. *See id* at 811. Obviously, the choice of these images could be debated.

217. The negative Black images included O.J. Simpson, Mike Tyson, Louis Farrakhan, Marion Barry, Arthur Washington, Lonny Gray, Tyshawn Williams, Charles Brackett, Michael McClinton, and Stanley Obas. Positive White images included Clint Eastwood, Jim Carrey, Tom Cruise, David Duchovny, Tom Hanks, Jay Leno, John F. Kennedy, Robert Redford, Norman Schwarzkopf, and Peter Jennings. *See id.* at 812.

218. The net decrease came from faster reaction times for the "Black + pleasant" and the "White + unpleasant" combinations in the IAT. Interestingly, the latencies for the "White + pleasant" and the "Black + unpleasant" combinations did not change across the various exemplar conditions. *See id.* at 807.

219. *Id.*

220. Isabelle Klein et al., *Retinotopic Organization of Visual Mental Images as Revealed By Functional Magnetic Resonance Imaging*, 22 COGNITIVE BRAIN RES. 26, 28-30 (2004).

221. 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, 828-29 (2001).

imagining a strong woman, her attributes and abilities, and the hobbies she enjoys; another group was instructed to imagine a Caribbean vacation.[222] Those who imagined the strong woman registered a significantly lower level of implicit stereotype in the IAT.[223] Based on follow-up experiments that included imagining different things and employed different measures of implicit bias than the IAT, the researchers concluded that the counter-stereotypic mental imagery caused the decrease in implicit stereotypes.[224]

*Countertypical teachers.* There is even evidence of external validity from real-world measurements. Nilanjana Dasgupta and Shaki Asgari tracked longitudinally female students before and after their first year of college.[225] Half the participants attended a coeducational college, whereas the other half attended a women's college. Both groups took tests measuring explicit and implicit bias against women and completed campus-experience questionnaires.[226] The two groups started with statistically indistinguishable levels of *implicit* bias: both groups viewed women stereotypically, as more "supportive" than "agentic."[227] After one year of college, however, the average implicit bias of the group that had attended women's colleges disappeared, whereas the implicit bias of the group that had attended coeducational colleges increased.[228] The researchers regressed campus environmental variables. For example, they asked whether lower implicit bias correlated with the number of courses taken with gender-related content, for example, in the Women's Studies department? The answer was no. The only statistically significant correlation was "exposure of female faculty"—defined as the number of women faculty and senior administrators these students encountered.[229]

These ISC findings suggest that debiasing does not have to take place solely through conventional peer-to-peer social contact, which is the mechanism emphasized in the contact hypothesis. Debiasing can also take place through repeat exposure to countertypical exemplars in positions of

---

222. *Id.* at 830.

223. *Id.* at 831. For the neutral imagery group, the reaction time difference between the schema-consistent and schema-inconsistent blocks was ninety-five milliseconds. For the counterstereotypic imagery group, the difference was twenty-four milliseconds, which reached statistical significance at $p < 0.05$. *See id.* at 831 tbl.1.

224. *Id.* at 837.

225. *See generally* 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 (2004).

226. *See id.* at 649-50.

227. *See* id. at 651.

228. The mean IAT effect for those attending a women's college started at 31 milliseconds and went down to -5 milliseconds. By contrast, the IAT effect for those attending a coed college started at 74 milliseconds and went up to 128 milliseconds. *Id.*

229. *Id.* ($p = 0.004$).

authority.[230] Furthermore, both the countertypical-visualizations and the countertypical-teachers studies demonstrated changes in implicit *stereotypes* not just attitudes. To the extent that peer-to-peer social contact is more effective in moderating attitudes than stereotypes, counterstereotypic exemplars may be an important way to moderate implicit stereotypes. These are not mutually exclusive mechanisms. Indeed, together, they may well help produce a larger culture conducive to debiasing.

### C. *Better Debiasing*

If images we see and imagine can decrease our implicit bias, an interesting range of possibilities become available for private, individual, voluntary, "do it yourself" attitude makeovers. How do you decorate your room? What is on your screensaver?[231] What is the office's décor?[232] We do not yet have definitive evidence that provides an uncontroversial list of best practices. But we do want to highlight the increasing evidence. For instance, Akalis and colleagues have begun a line of research showing that it is possible through the ordinary concentration on particular thoughts (positive thoughts about those who are overweight, for example) to reduce bias on measures such as the IAT. Moreover, they show that compared to a control group, yoga practitioners in India can lower their IAT bias after five minutes of concentration. Notably, the yoga practitioners do not differ from the control group at baseline - they show the same quite high bias favoring their ingroup whether it is Indians (compared to Pakistanis), Hindus (compared to Muslims) or high castes (compared to low castes).[233] These findings suggest that mental exercises might provide a path, however little understood at this time, which establishes control over seemingly uncontrollable attitudes.

For purposes of revising affirmative action discourse, we focus not on private, individual, voluntary makeovers, which do not raise the most difficult legal and policy questions. What if, instead, drawing on results such as Dasgupta and Asgari's study, an institution hires certain people because of their debiasing capacity on their students, customers, or employees? After all, Brian Lowery, et al. showed that the mere presence of an African

---

230. *Cf.* Jonathan Alger, *When Color-Blind is Color-Bland: Ensuring Faculty Diversity in Higher Education*, 10 STAN. L. & POL'Y REV. 191, 195 (1999) (suggesting that faculty diversity may be more important than student diversity in decreasing stereotypes and observing that this is "an entirely different sort" of role model theory).

231. *See* Kang, *supra* note 12, at 1537.

232. *See, e.g.*, Jolls & Sunstein, *supra* note 12, at Part. II.B.3 (discussing manipulation of various environmental stimuli, such as portraits on the wall).

233. Scott A. Akalis, Jhansi Nannapaneni, & Mahzarin R. Banaji, Do-It-Yourself Mental Makeovers: How Directed Thinking Influences Implicit Attitudes (2006) (unpublished manuscript, on file with Harvard University).

American experimenter reduced race bias in White participants.[234] Debiasing worked presumably because participants heard the African American instructor give directions, be in charge, and implicitly hold power. Likewise the studies discussed in the prior section provide us further reason to think that such findings will appear in the scientific literature shortly.[235] What would be the legal and policy implications of hiring what we call "debiasing agents"?

### *1. Debiasing Agents*

A debiasing agent is an individual with characteristics that run counter to the attitudes and/or the stereotypes associated with the category to which the agent belongs. Examples include women construction workers, male nurses, Black intellectuals, White janitors, Asian CEOs, gay boxers, and elderly marathon runners. In our times, individuals such as Toni Morrison, Lance Armstrong, Tiger Woods, and Condoleezza Rice are debiasing agents. Mahatma Gandhi and Martin Luther King, Jr. were debiasing agents in their time. In order for the debiasing agent to be successful, perceivers must map the agent to the relevant social category. Further, perceivers must not dismiss the agent as a mere exception, but must recognize the agent as anchored to that social category, notwithstanding his or her countertypical qualities.

Although similar, we disentangle "debiasing agent" from the more familiar "role model." The traditional role-model argument suggests that we should grant affirmative action to women and racial minorities in teaching and leadership positions because they can act as "models" for people of the same social category. Both the Left and Right have protested the role-model argument. For example, Richard Delgado argues that being a role "model" requires women and minorities to be assimilationist, to suffer burdens not placed on Whites and males, and to perpetuate a system-reinforcing meritocratic myth that if you work hard, you can succeed "just like me."[236] At the same time, Justice Thomas chafes at the assumption that

---

234. *See* Brian S. Lowery, et al., *Social Influence Effects on Automatic Racial Prejudice*, 81 J. PERSONALITY & SOC. PSYCHOL. 842 (2001).

235. *Cf.* Jolls & Sunstein, *supra* note 12, at Part II.B.2 (recommending diversity in the supervisory workforce); *see also* Stacey Sinclair et al., *Social Tuning of Automatic Racial Attitudes: The Role of Affiliative Motivation*, 89 J. PERSONALITY & SOC. PSYCHOL. 583, 590 (2005) (finding evidence in support of an "affiliative social tuning hypothesis" in which White participants' implicit biases aligned with the ostensible attitudes of experimenters that participants liked).

236. Richard Delgado, *Affirmative Action as a Majoritarian Device: Or, Do You Really Want to Be a Role Model?*, 89 MICH. L. REV. 1222, 1226-29 (1991). Various commentators, in light of this criticism, have tried to reframe the idea of "role model" to being a "mentor" or a "connected critic." *See, e.g.*, Enrique R. Carrasco, *Collective Recognition as a Communitarian Device: Or, Of Course We Want to Be Role Models!*, 9 LA RAZA L.J. 81, 94-96 (1996) ("connected critic" drawing on the work of philosopher Michael Walzer); Lani Guinier, *Of Gentlemen and Role Models*, 6 BERKELEY WOMEN'S L.J. 93, 100-05 (1991) (mentor); Taunya Lovell Banks, *Two Life Stories: Reflections of One Black Woman Law Professor*, 6 BERKELEY WOMEN'S L.J. 46, 46 (1991) (mentor).

we can only look up to those of the same race or gender.[237] He would reject the balkanizing notion that an Asian student can and should look up to an Asian professor exclusively.

Courts have also raised legal objections. Under an equal protection analysis, the Supreme Court in *Wygant* rejected a role-model theory because it did not seek to remedy past acts of discrimination, but instead smacked of forward-looking social engineering, lacked a limitation principle that would circumscribe the scope and duration of the program, and encouraged a form of racial tracking between role model and beneficiary that the court found distasteful.[238] A general role-modeling justification has also been rejected under Title VII.[239]

The debiasing agent justification differs from the role-model argument in two ways. First, the beneficiary class is not restricted to those students who occupy the same social category as the debiasing agent. Instead, debiasing affects all students and arguably is most important for those who are not in the same social category. This difference avoids much of the balkanization critique described above. Second, the objective in employing debiasing agents is not to increase students' self-esteem or self-confidence, or to catalyze their ability to imagine a different future, which may seem like an open-ended, forward-looking social engineering objective. (If this is a secondary effect, so be it, but that is not the purpose of debiasing agents.) Instead, the purpose of the debiasing agent is to mitigate objectively measurable bias by producing environmental conditions that alter the strength of association between social category and attitude or attribute. This fits squarely within the presentist framing of reducing discrimination here and now.

---

237. Justice Thomas has critiqued the Court's voting dilution cases for resting on the "implicit assumption that members of racial and ethnic groups must all think alike on important matters of public policy and must have their own 'minority preferred' representatives . . . if they are to be considered represented at all." Holder v. Hall, 512 U.S. 874, 903 (1994) (Thomas, J., concurring). He would argue that the role-model theory perpetuates the divisive notion that "members of [a] racial group must think alike and [have] their [own] interests . . . distinct [from other racial groups]." *Id.* at 906. Similarly, Shelby Steele argues that race-based interventions such as "the ubiquitous idea of racial role models" act to "suppress black individuals with the mark of race just as certainly as segregation did, by relentlessly telling them that their racial identity is the most important thing about them." SHELBY STEELE, A DREAM DEFERRED: THE SECOND BETRAYAL OF BLACK FREEDOM IN AMERICA 61 (1998).

238. *See* Wygant v. Jackson Bd. of Educ., 476 U.S. 267 (1986). In her concurring opinion, Justice O'Connor distinguished diversity from the role model rationale. *See id.* at 289 ("The goal of providing 'role models' . . . should not be confused with the very different goal of promoting racial diversity among the faculty."). Justice Marshall emphasized that the beneficiary class of diversity is the entire student body. *See id.* at 306 (Marshall, J., dissenting) (recommending remanding the case and noting that if it could be established that the hiring policy "sought to achieve diversity and stability for the benefit of *all students*" it would be constitutional) (emphasis in original).

239. *See* Taxman v. Piscataway Twp. Bd. of Educ., 91 F.3d 1547 (3d Cir. 1996) (en banc). Further discussion of *Taxman* appears *infra* text accompanying note 254.

*2. Legality*

Consider the following hypothetical: A business school hires an Asian professor partly to decrease bias against Asians[240] among business school students. A White candidate with comparable qualifications who did not get hired sues the (public) business school. What are the merits of the White person's equal protection and Title VII claims?

*Equal protection.* Given *Wygant*'s rejection of the "role model" argument,[241] could a debiasing agent justification nevertheless withstand equal protection scrutiny? First, we must determine the appropriate standard of review. One could plausibly argue that crediting individuals for their debiasing capacities is not a facial racial classification and thus does not warrant strict scrutiny. That is because a particular race is neither necessary nor sufficient for debiasing. For example, a person who does not regard herself as Asian (perhaps because she has an interracial background and regards herself as *hapa*) may nevertheless decrease anti-Asian implicit bias (because she "looks" Asian).[242] Conversely, not all Asians will have debiasing capacity; only those who overperform countertypical qualities will. Even if this formalistic argument is plausible,[243] we do not pursue it. We are interested in tackling the hardest case, and thus assume that strict scrutiny applies.

That means we must demonstrate a "compelling" interest. Again, the goal of debiasing is not to increase self-esteem, but to decrease discrimination caused by implicit bias. As we demonstrated in Part I, that interest is "compelling." One distinction here is that the goal is not to prevent discrimination by the business school itself but by its future graduates. Since there is some uncertainty about the degree to which any actor can try to respond to discrimination committed by another actor, this distinction is worth further examination. In the end, however, this distinction should

240. Asians are viewed as competent but not sociable. *See* Monica H. Lin et al., *Stereotype Content Model Explains Prejudice for an Envied Outgroup: Scale of Anti-Asian American Stereotypes*, 31 PERSONALITY & SOC. PSYCHOL. BULL. 34 (2005) (explicit self-reports). Also, Whites are viewed as more "American" than "Foreign" as compared to Asian Americans. *See* Thierry Devos & Mahzarin R. Banaji, *American = White?*, 88 J. PERSONALITY & SOC. PSYCHOL. 447, 453, 457 (2005) (using IAT). For narrative discussion of anti-Asian bias in the law reviews, see, e.g., Keith Aoki, *"Foreign-ness" & Asian American Identities: Yellowface, World War II Propaganda and Bifurcated Racial Stereotypes*, 4 ASIAN PAC. AM. L.J. 1 (1999); Jerry Kang, *Racial Violence Against Asian Americans*, 106 HARV. L. REV. 1926 (1993).

241. *See Wygant*, 476 U.S. at 274-76; *see also* City of Richmond v. J.A. Croson Co., 488 U.S. 469, 497-98 (plurality opinion) (citing and summarizing *Wygant*'s view of role models).

242. Suppose that Tiger Woods does not regard himself as Black (his father self-identifies as African American; his mother self-identifies as Thai.) Nonetheless, he may decrease implicit bias against Blacks.

243. *See, e.g.*, Jack M. Balkin & Reva B. Siegel, *The American Civil Rights Tradition: Anticlassification or Antisubordination?*, 58 U. MIAMI L. REV. 9, 14-20 (2003) (demonstrating indeterminacy of what counts as a racial classification).

make no constitutional difference, in light of the Supreme Court's decision in *Grutter*.

In *Grutter v. Bollinger*, the Court emphasized that student diversity was valuable because it could help "break down racial stereotypes."[244] The same can be said about debiasing agents. The Court also stressed the value of training future workers and leaders to operate in an "increasingly diverse workforce and society."[245] Although the Court did not specify a psychological mechanism, it implied that through exposure and interaction with diverse others, students will learn to cooperate better across social categories and gain skills that will be increasingly valuable for business success and national security. Again, the same goes for debiasing agents, except that the psychological mechanism is more sharply delineated. If the Court deems student diversity sufficiently "compelling" for miscellaneous pedagogical reasons including its tendency to decrease bias in future leaders, workers, and citizens, then it should find the more focused objective of decreasing implicit bias even more compelling.

In addition to serving a compelling state interest, an employer's consideration of a candidate's debiasing capacity must be "narrowly tailored." In this analysis, courts consider various factors, such as who benefits and by how much, who is harmed and by how much, and whether the program seems to be justified and bounded by something more than intuition. A flexible "plus" factor, allowing a small, but not decisive, preference for debiasing capacity would satisfy the requirements of *Grutter*. Those burdened would simply have a slightly smaller chance of being hired; they would not lose a vested interest such as a job or pension. Finally, the entire program would be justified by science more rigorous than that accepted by the *Grutter* Court.

If this sounds implausible, consider the line of "operational needs" cases that have allowed race-based hiring in the law enforcement and corrections context.[246] Judge Richard Posner's opinion in *Wittmer v. Peters*[247] is instructive. That case involved a "boot camp" that served as an alternative to conventional prison. Sixty-eight percent of the inmates were Black.[248] The boot camp hired a Black applicant as lieutenant over higher-testing Whites. The Seventh Circuit Court of Appeals upheld the constitutionality of that decision.

A classroom may not seem comparable to boot camp (although some might disagree). But there are interesting analogies. As Judge Posner explained, the justification for preferring the Black lieutenant in boot camp

244. Grutter v. Bollinger, 539 U.S. 306, 330 (2003).

245. *Id.* at 321.

246. *See, e.g.*, Wittmer v. Peters, 87 F.3d 916, 920 (7th Cir. 1996); Barhold v. Rodriguez, 863 F.2d 233, 238 (2d Cir. 1988); Detroit Police Officers' Ass'n v. Young, 608 F.2d 671, 695 (6th Cir. 1979).

247. 87 F.3d 916 (7th Cir. 1996) (Posner, J.).

248. *See id.* at 917.

was not a traditional role-model argument: the point was not to encourage the inmates to become prison guards.[249] The point also was not to create proportional racial balancing between the staff and the inmates.[250] Instead, "[t]he black lieutenant is needed because the black inmates are believed unlikely to play the correctional game of brutal drill sergeant and brutalized recruit unless there are some blacks in authority in the camp."[251] Similarly, the point of debiasing agents is not to encourage Asians to become professors (although, again, one would not object to that secondary effect). Rather, it is to decrease the implicit bias within the entire student body.

Arguably the case for debiasing agents is even stronger than the case for the Black lieutenant in *Wittmer*. A colorblind enthusiast could argue that Black inmates' demand for Black lieutenants is a sort of pass-through discrimination that should not be tolerated. After all, we do not let White customer preference for White hostesses justify an employer's preference for White hostesses, so why should we allow such a justification in the boot camp context?

This objection, however, simply does not apply to the debiasing agent rationale. The objection might have traction if a business school were hiring an Asian professor to teach accounting because it speculated that its students saw greater credibility or legitimacy in having a person of Asian descent teach those classes.[252] However, a debiasing agent is not deployed to *respect* color or gender-conscious preferences within the student body; it is targeted to *override* them. A debiasing agent is not sought to secure some amorphous benefit such as "institutional legitimacy," within a forward-looking frame. Instead, it is relentlessly focused on preventing discrimination by decreasing demonstrable levels of bias among those within its institution.

---

249. *See id.* at 920; *see also* Barhold v. Rodriguez, 863 F.2d 233, 238 (2d Cir. 1988) ("Operational need" refers to a law enforcement body's need to carry out its mission effectively, with a workforce that appears unbiased, is able to communicate with the public and is respected by the community it serves. "Role models," in contrast, are people whose very existence conveys a feeling of possibility to others; they give hope that a previously restricted opportunity might now be available.).

250. *Wittmer*, 87 F.3d at 920.

251. *Id.* This type of reasoning is similar to some of the language in *Grutter*. *See, e.g.*, *Grutter*, 539 U.S. at 332 ("In order to cultivate a set of leaders with legitimacy in the eyes of the citizenry, it is necessary that the path to leadership be visibly open to talented and qualified individuals of every race and ethnicity. All members of our heterogeneous society must have confidence in the openness and integrity of the educational institutions that provide this training.").

252. *Cf.* Ferrill v. Parker Group, Inc., 168 F.3d 468, 475 (11th Cir. 1999) (finding illegal under 42 U.S.C. § 1981 an employer's policy of having Black employees call Black voters because it was based on stereotypes, and rejecting the idea that race was a bona fide occupational qualification); Knight v. Nassau County Civil Serv. Comm'n, 649 F.2d 157, 162 (2d Cir. 1981), *cert denied*, 454 U.S. 818 (1981) (holding that assigning an employee to do minority recruitment because of his race violates Title VII).

The case for debiasing agents also rests on a firmer scientific foundation than did *Wittmer*. In *Wittmer*, Judge Posner emphasized that expert testimony supported the "operational needs" justification.[253] He also noted that some flexibility should be granted for experimentation, and that after more research is done, the constitutional balance could be reconsidered.[254] But on the record presented to the court, the racial preference was not unconstitutional. Again, the same goes for debiasing agents. The scientific foundation for debiasing agents is at least as strong as that deemed acceptable in *Wittmer*. And, as the science improves, debiasing agent justifications and implementations can be adjusted accordingly.

*Title VII.* Even if debiasing-agent justifications pass the equal protection test, do they fail under Title VII, which may reject non-remedial objectives? In *Taxman v. Piscataway Township Board of Education*,[255] the Third Circuit Court of Appeals en banc held that "a non-remedial affirmative action plan, even one with a laudable purpose, cannot pass [Title VII's] muster."[256] The case involved a school board's decision to break a tie between a White and Black faculty member in a layoff decision in favor of the Black teacher for forward-looking purposes.[257] Interpreting the Supreme Court's precedents *Weber* and *Johnson*, the *Taxman* court explained that in order for a voluntary affirmative-action plan to be lawful, it must "have purposes that mirror those of the [Title VII] statute".[258] As a matter of statutory interpretation, the court held that Title VII was enacted to further two goals only: to end discrimination, and to remedy the consequences of past discrimination.[259] Since the Piscataway school board's justification did not fit these objectives, its decision to break the tie on the basis of race violated Title VII. No "additional non-remedial deviations"[260] would be tolerated.

There are, of course, good arguments on the other side; *Taxman* itself generated four separate dissenting opinions. Further, after the reaffirmation of diversity in *Grutter*, the practical precedential value of *Taxman* may have weakened—even though *Grutter* was about equal protection, not Title VII. In any event, the preference given to debiasing agents satisfies even *Taxman*'s demanding standards. Again, the goal of crediting debiasing capacity is not to further some general value in "diversity." Rather, it is to stop discrimination by decreasing the implicit bias in students, who will graduate to become future workers, employers, and leaders.

253. *Wittmer*, 87 F.3d at 920. Two years later, in McNamara v. City of Chicago, 138 F.3d 1219 (7th Cir. 1998), the Seventh Circuit declined to extend the reasoning of *Wittmer* to firefighters because the evidentiary record did not exist. *See id.* at 1222.

254. *Wittmer*, 87 F.3d at 920-21.

255. 91 F.3d 1547 (3d Cir. 1996) (en banc).

256. *Id.* at 1550.

257. *Id.* at 1551-52.

258. *Id.* at 1550 (quoting United Steelworkers v. Weber, 443 U.S. 193 (1979)).

259. *Id.* at 1557.

260. *Id.* at 1558.

Constitutionally, this is a compelling interest. Statutorily, this objective is consonant with the goals of Title VII, even as narrowly interpreted in *Taxman*.[261]

In sum, ISC outlines a new mechanism for decreasing implicit bias: debiasing agents. We have not addressed various complications. For instance, if the attitude toward a category is negative, then a friendly, accommodating individual may be more effective in changing that attitude. This would favor a particular sort of identity performance, one that provides racial comfort, not confrontation.[262] These complications, although significant, are not insurmountable. An even more powerful version of this critique applies to the identity performance demanded of "role models." Nevertheless, employers and educators regularly favor candidates who can also be role models without theoretical angst. We also repeat that debiasing agents are chosen not for their stereotypical qualities but for their counter-typical ones. In the end, we have made a plausible case that preferring certain individuals because of their debiasing capacity can be a lawful fair measure, especially in the context of higher education. The rationale is not self-esteem; it is disinfection.

## CONCLUSION
## A NEW ENDING

So, when should affirmative action end? This is a vexing question even for many proponents of affirmative action. Those who view affirmative action as having backward-looking justification seek res judicata on corrective justice claims. Those who view affirmative action as having forward-looking justifications wonder whether there is any end point at all. Indeed opponents of affirmative action regularly ask this question as a debater's ploy to elicit a vague answer, which is then proffered as dispositive evidence that affirmative action advocates are indefinite social engineers.

Recently, Justice O'Connor in her *Grutter* opinion predicted twenty-five years.[263] A rich literature interpreting what that deadline might mean as

261. Title VII also requires that voluntary affirmative action programs not "unnecessarily trammel the interests of the [nonminorities]." *Weber*, 443 U.S. at 208. Those attributes that get the debiasing agent program past equal protection "narrow tailoring" analysis simultaneously satisfy these Title VII requirements.

262. *See, e.g.*, Carbado & Gulati, *supra* note 37, at 1805-06 (discussing "racial comfort"). Recall that the Dasgupta and Asgari study placed Cosby and Farrakhan on different sides of the attitudinal divide. This is where cultural studies meets cognitive psychology. As another example, debiasing ability may become an additional job requirement implicitly placed on people of color and women but not on White males. If there is only one slot for a woman or minority (given limited resources and tokenism), an institution may decline to hire a candidate without demonstrable debiasing capacity even though they do not require that a similar question is asked of White male candidates. Accordingly, a soft spoken Asian woman may be penalized, whereas a soft spoken White man may not.

263. Grutter v. Bollinger, 539 U.S. 306, 343 (2003) ("It has been 25 years since Justice Powell first approved the use of race to further an interest in student body diversity in the context of public higher education. Since that time, the number of minority applicants with high grades and test scores

a matter of politics and precedent has already appeared. Many commentators observe that the deadline is mechanical, arbitrary,[264] and arguably incoherent given the "diversity" justifications that O'Connor accepted. After all, if racial integration produces pedagogical advantages through "diversity," why should those benefits evaporate twenty-five years from now? Notwithstanding such complaints, it takes a theory to beat a theory, and critics are hard pressed to provide an alternative that is, on the one hand, specific and objectively determined, and, on the other hand, less arbitrary than O'Connor's deadline.

Implicit social cognition (ISC) offers new alternatives. The terminus question can be framed at two levels of generality. At a specific level, the question might be about when any specific fair measure becomes no longer necessary. That question is conceptually easy to answer and depends on why the specific fair measure was adopted in the first place. For example, steps taken to provide fairer measures of merit can sunset when we demonstrate that mismeasurement is no longer taking place. The narrow tailoring here is obvious.

The more interesting question is framed at a higher level of abstraction. Generally speaking, when should "affirmative action" policies end? We offer the following terminus: Fair measures that are race- or gender-conscious[265] will become presumptively unnecessary when the nation's implicit bias against those social categories goes to zero or its negligible behavioral equivalent.[266] For all those who praise colorblindness, this will be when we as a nation become truly colorblind, not only to visible light but also to the infrared frequencies that lurk beneath.[267]

This terminus rejects hypocrisy and self-deception. Recall the fascinating finding of dissociation—that explicit self-reports of bias do not line up with implicit measures of bias. Each of us has commented, in different contexts, about the value of consciously struggling to rein in implicit bias that is inconsistent with our explicit normative commitments.[268] This

---

has indeed increased. We expect that 25 years from now, the use of racial preferences will no longer be necessary to further the interest approved today.") (citation omitted).

264. *See, e.g.*, Spann, *supra* note 24, at 248 n.123.

265. Fair measures that are not race- or gender-conscious and that improve the accuracy of measurements would never have to end.

266. What would count as a "negligible equivalent" is related to the strength of linkage between implicit bias and real-world behavior, which is what we are most interested in. If implicit bias has been reduced sufficiently such that behavior is not being affected, we could consider that level to be negligible.

267. Justice Blackmun once wrote that "In order to get beyond racism, we must first take account of race." Regents of Univ. of Cal. v. Bakke, 438 U.S. 265, 407 (1978) (Blackmun, J., concurring in part and dissenting in part). The ISC variant of that paradox is that "In order to be blind, we must first see the invisible."

268. *See* Mahzarin R. Banaji, *The Opposite of a Great Truth Is Also True: Homage to Koan #7*, *in* PERSPECTIVISM IN SOCIAL PSYCHOLOGY: THE YIN AND YANG OF SCIENTIFIC PROGRESS 127, 134 (John T. Jost et al. eds., 2003) ("[O]ne measure of the evolution of a society may indeed be the degree of

terminus point demands alignment between our explicit self-descriptions and our implicit states of mind.[269] Verbal gestures are not enough. Honest mistakes, although honest, are still mistakes.

Also, this terminus has the advantage of being nonarbitrary, in sharp contrast to Justice O'Connor's suggestion of twenty-five years, which Justice Thomas read as a statute of limitations more than a mere prediction.[270] Moreover, this terminus is specific and objectively measurable. Implementing a national gauge may sound implausible, but it is not wishful thinking. Modern technology can already measure where an entire society stands on various biases through reliable and anonymous web-based data collection[271] To be sure, our proposal triggers myriad questions about which specific measures of implicit bias, taken through which specific instruments, against which specific categories, with which specific level of confidence. Still, compared to other proposals on the table, we believe that this one has the potential to be the most completely specified, objectively implementable, and narrowly tailored.

Our revision of "affirmative action" into "fair measures" provides a new framing for social intervention in favor of equality. Unwarranted discrimination exists here and now: it can be documented through scientific methods that cannot be dismissed as hyperbole or playing the "race card." We are not merely pointing to disparities between social groups, which in and of itself may not trouble many Americans. Instead, we suggest that some of these disparities may be caused by discrimination that arises from ordinary forms of implicit bias in our minds, here and now.

The very same science that allows us to document implicit bias and its discriminatory consequences also provides new insight into the ways that we can individually and collectively debias our institutions and ourselves. Taking fair measures is simply the implementation of these measures. These measures will not, and are not designed to, respond to all genuine and meritorious claims on the distribution of scarce resources. As explained earlier, corrective and distributive justice claims might be made under the banner of "affirmative action" or "social justice" that "fair

---

separation between conscious and unconscious attitudes—that is, the degree to which primitive implicit evaluations that disfavor certain social groups or outgroups are explicitly corrected at the conscious level at which control is possible."); Kang, *supra* note 12, at 1587-88 ("[S]ocial strategies that decrease the dissociation that we as a society collectively experience should be seen as autonomy-reinforcing . . . . How could it be against autonomy to bring our implicit thoughts in line with our explicit ones?").

269. This argument takes advantage of the contingent fact that our official public ideology rejects bias across social categories. In other words, if we lived in a social and legal system that officially endorsed castes, in which explicit bias against lower social categories was accepted, we would not call for alignment as the solution.

270. Justice Thomas took this deadline seriously and wrote "that racial discrimination in higher education admissions will be illegal in 25 years." *Grutter*, 539 U.S. at 351 (Thomas, J., concurring in part and dissenting in part).

271. *See* Nosek, *supra* note 47.

measures" simply does not cover. We neither foreclose nor dismiss those conversations and struggles. But to these crucial ongoing debates, we seek to add something new, something that science has uncovered. Driven by a behavioral realist accounting of implicit social cognition, we sketch out a new terrain of fair measures. Although not vast, it is a significant place, one that establishes the foundation for tie-breakers and debiasing agents while demolishing facile assurances that we are all already colorblind. This terrain, we hope, can become new common ground—a field that seeks to be level, measured, and fair.