

ARTICLES

(HOW) DOES UNCONSCIOUS BIAS MATTER?: LAW, POLITICS, AND RACIAL INEQUALITY

*Ralph Richard Banks**
*Richard Thompson Ford***

ABSTRACT

During the past several years, psychological research on unconscious racial bias has grabbed headlines, as well as the attention of legal scholars. The most well-known test of unconscious bias is the Implicit Association Test (IAT), a sophisticated and methodologically rigorous computer-administered measure that has been taken by millions of people and featured in major media. Its proponents contend that the IAT reveals widespread unconscious bias against African Americans, even among individuals who believe themselves to be free of racial bias.

In fact, however, the findings of the IAT are ambiguous. The test could just as plausibly be thought to measure racial bias that is simply covert, known to oneself yet intentionally concealed from researchers. On this interpretation, the IAT reveals not that individuals are more biased than they realize, but that they are more biased than they want others to know. The characterization of

* Jackson Eli Reynolds Professor of Law, Stanford Law School.

** George E. Osborne Professor of Law, Stanford Law School.

the IAT as a measure of unconscious bias has practically eclipsed this plausible alternative interpretation. Why?

One possibility is that unconscious bias, even if not incontrovertibly demonstrated by the IAT, warrants attention because it poses a unique challenge for antidiscrimination doctrine. But this explanation for the ascendance of the unconscious bias discourse is wrong. Antidiscrimination law grapples as well, or as poorly, with unconscious bias as with covert bias. Neither statutory nor constitutional doctrine turns on the distinction between the two.

The better explanation for the ascendance of the unconscious bias discourse is that assertions of widespread unconscious bias are more politically palatable than parallel claims about covert bias. The invocation of unconscious bias levels neither accusation nor blame, so much as it identifies a quasi-medical ailment that distorts thinking and behavior. People may be willing to acknowledge the possibility of unconscious bias within themselves, even as they would vigorously deny harboring conscious bias. The unconscious bias claim thus facilitates a consensus that the race problem persists.

Despite its ostensible political benefits, the unconscious bias discourse is as likely to subvert as to further the cause of racial justice. Racial injustice inheres in the entrenched substantive racial inequalities that pervade our society. These disparities are not primarily a consequence of contemporary racial bias. Thus, the goal of racial justice efforts should be the alleviation of substantive inequalities, not the eradication of unconscious bias. Yet, the rhetoric of unconscious bias is so compelling that people are likely to accept it as the goal of racial reform and, consequently, to push the theory in directions that siphon energy away from problems of substantive inequality and that may be undesirable in their own right. The unconscious bias discourse reinforces a misguided preoccupation with mental state, and perpetuates an obsession with antidiscrimination law, rather than policy reform, as a means of realizing racial justice goals. If the goal is to eliminate substantive inequalities, then the task of racial justice advocates should be to explain forthrightly why those inequalities are objectionable and how to address them.

INTRODUCTION	1056
I. EMPIRICAL AMBIGUITY	1060
A. <i>The Measurement of Implicit Bias</i>	1060
B. <i>The Necessity of Measuring Conscious Bias</i>	1063
C. <i>The Social Desirability Problem</i>	1065
D. <i>Recharacterizing the IAT</i>	1068
II. ANTIDISCRIMINATION DOCTRINE	1072
A. <i>Statutory Law: Title VII</i>	1073
1. <i>Single-Motive Cases: Formalism, Realism, and Employment at Will</i>	1073
2. <i>Mixed-Motive Cases: Causation and Duty</i>	1081
3. <i>The Honest-Belief Rule</i>	1086
B. <i>Constitutional Law: Equal Protection</i>	1089
C. <i>Disparate Impact, Affirmative Action, and the Burden of Proof</i>	1100
III. POLITICAL APPEAL	1103
A. <i>The Stigma of Racism</i>	1103
B. <i>The Denial of Discrimination</i>	1104
C. <i>The Denial of Bias</i>	1106
D. <i>The Historical Narrative</i>	1108
IV. AGAINST BIAS	1110
A. <i>The IAT and the Risk of Failure</i>	1110
1. <i>Empirical Uncertainty</i>	1110
2. <i>The Scholarly Role</i>	1112
B. <i>The Costs of Success</i>	1113
1. <i>Goal Distortion: Confusion of Means and Ends</i>	1113
2. <i>Undesirable Outgrowths</i>	1116
CONCLUSION	1121

INTRODUCTION

Jimmy Carter made headlines when he confessed—not to his pastor, psychotherapist, or spouse, but to a national audience—that he had “looked on a lot of women with lust.”¹ It may sound odd to claim that the 1970s were a more innocent era but let’s face it: these days it’s enough if a politician doesn’t sin in the Oval Office, with an underage congressional page, or using a .gov e-mail address. Virtually no one, let alone a politician, would admit to sinning in one’s own heart—the one place on earth no special prosecutor can serve a summons and no intrepid reporter can plant a tape recorder. Yet lots of people want to look just there—in one’s own heart of hearts and innermost mind—for sin.

Repressed memories, subconscious phobias, the maledictions of the reactive mind—popular psychology is replete with hypotheses that posit that reality is hidden, that appearances deceive, that the truth is submerged, obscure, invisible to the naked eye. The idea of the unconscious is highly provocative, and appealing, perhaps nowhere more so than as dramatized in Hitchcock’s series of psychodramas (most famously *Psycho*, but also *Vertigo*, *Spellbound*, and *Marnie*).² The idea of the unconscious has proven appealing as well for legal scholars writing about antidiscrimination law during the past two decades. Early legal analyses drew from the theories of Freud, for whom dreams and free association served as passageways to the depths of the unconscious.³ More recently, legal scholars have relied on a growing body of sophisticated and methodologically rigorous experimental studies conducted by social and cognitive psychologists.⁴

¹ Robert Scheer, *Playboy Interview: Jimmy Carter*, PLAYBOY, Nov. 1976, at 86.

² See PSYCHO (Shamley Productions 1960); VERTIGO (Paramount Pictures 1958); SPELLBOUND (Vanguard Films 1945); MARNIE (Universal Pictures 1964).

³ See Charles R. Lawrence III, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317 (1987).

⁴ The psychological literature generally refers to implicit bias. Throughout this Article, we use the terms “implicit” and “unconscious” interchangeably. The legal literature is large and growing. See, e.g., Samuel R. Bagenstos, *Implicit Bias, “Science,” and Antidiscrimination Law*, 1 HARV. L. & POL’Y REV. 477 (2007); Samuel R. Bagenstos, *The Structural Turn and the Limits of Antidiscrimination Law*, 94 CAL. L. REV. 1 (2006); R. Richard Banks, Jennifer L. Eberhardt & Lee Ross, *Discrimination and Implicit Bias in a Racially Unequal Society*, 94 CAL. L. REV. 1169 (2006); Gary Blasi, *Advocacy Against the Stereotype: Lessons from Cognitive Social Psychology*, 49 UCLA L. REV. 1241 (2002); Theodore Eisenberg & Sheri Lynn Johnson, *Implicit Racial Attitudes of Death Penalty Lawyers*, 53 DEPAUL L. REV. 1539 (2004); Tristin K. Green, *A Structural Approach as Antidiscrimination Mandate: Locating Employer Wrong*, 60 VAND. L. REV. 849 (2007); Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CAL. L. REV. 945 (2006); Olatunde C.A. Johnson, *Disparity Rules*, 107 COLUM. L. REV. 374 (2007); Christine Jolls &

The most well-known test of unconscious bias, and the one typically referenced by legal scholars, is the Implicit Association Test (IAT).⁵ A computer-administered test available over the Internet, the IAT is a compelling interactive experience that has been taken by millions of people, and featured in print and broadcast media.⁶ The IAT measures the strength of the association between social categories (e.g., blacks or whites) and positive and negative attributes (e.g., “joy” and “love” versus “agony” and “evil”). Akin to a computer game for grownups, the IAT requires momentary immersion into the interactive medium. In a series of trials, the participant categorizes images or words that appear on the computer screen by pressing particular computer keyboard keys as quickly as possible. At the end of the exercise, the computer calculates a score that reflects the nature and magnitude of one’s implicit bias.⁷ Most participants are found to have an implicit bias against African Americans. The overt racism of the Jim Crow era, the psychological research suggests, has given way to racial bias that is predominantly unconscious.

Cass R. Sunstein, *The Law of Implicit Bias*, 94 CAL. L. REV. 969 (2006); Jerry Kang, *Trojan Horses of Race*, 118 HARV. L. REV. 1489, 1510 (2005) [hereinafter Kang, *Trojan Horses of Race*]; Jerry Kang & Mahzarin R. Banaji, *Fair Measures: A Behavioral Realist Revision of “Affirmative Action,”* 94 CAL. L. REV. 1063 (2006); Linda Hamilton Krieger & Susan T. Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 CAL. L. REV. 997 (2006); Audrey J. Lee, *Unconscious Bias Theory in Employment Discrimination Litigation*, 40 HARV. C.R.-C.L. L. REV. 481 (2005); Justin D. Levinson, *Forgotten Racial Equality: Implicit Bias, Decisionmaking, and Misremembering*, 57 DUKE L.J. 345 (2007); Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 OHIO ST. L.J. 1023 (2006); Rachel F. Moran, *Whatever Happened to Racism?*, 79 ST. JOHN’S L. REV. 899 (2005); John A. Powell, *Structural Racism: Building upon the Insights of John Calmore*, 86 N.C. L. REV. 791 (2008); Russell K. Robinson, *Perceptual Segregation*, 108 COLUM. L. REV. 1093 (2008); Reshma M. Saujani, “*The Implicit Association Test*: A Measure of Unconscious Racism in Legislative Decision-Making,” 8 MICH. J. RACE & L. 395 (2003); Michael S. Shin, *Redressing Wounds: Finding a Legal Framework to Remedy Racial Disparities in Medical Care*, 90 CAL. L. REV. 2047 (2002).

⁵ Greenwald and Banaji have been the two central figures in the development and growing popularity of the IAT, which has been the subject of more than 300 legal and psychological articles. See, e.g., Anthony G. Greenwald et al., *Measuring Individual Difference in Implicit Cognition: The Implicit Association Test*, 74 J. PERSONALITY & SOC. PSYCHOL. 1464 (1998). The IAT has received substantial publicity in the popular press. See, e.g., Diane Cole, *Don’t Race to Judgment*, U.S. NEWS & WORLD REP., Dec. 26, 2005, at 90; Shankar Vedantam, *See No Bias*, WASH. POST, Jan. 23, 2005, (Magazine), at 12.

⁶ A Google search for “Implicit Association Test” in May, 2009 yielded 42,000 results. For a brief sampling of media accounts, see, for example, *How the IAT Works*, BOSTON GLOBE, Dec. 19, 2004, at K5; Courtland Milloy, *Out from Under the Thumb of White Bias*, WASH. POST, Jan. 26, 2005, at B1; Jay Dixit, *Screen Test: Why We Should Start Measuring Bias*, SLATE, Jan. 26, 2006, <http://www.slate.com/id/2134921/>; Joel Schwarz, *Study Suggests Polls May Overestimate Support for Obama, Underestimate Backing for Clinton Among Democrats*, U. WASH. NEWS, Dec. 18, 2007, <http://uwnews.org/article.asp?articleid=38613>; Vedantam, *supra* note 5; Lucy Wilkins, *Are You Racist? The Test That Claims To Know*, BBC NEWS, Apr. 20, 2005, <http://news.bbc.co.uk/1/hi/magazine/4447471.stm>.

⁷ Most people who take the Race IAT (many African Americans among them) are found to have an anti-black implicit bias.

In fact, the findings of the IAT are ambiguous. The characterization of the IAT as a measure of implicit bias depends on being able to distinguish implicit bias from conscious bias. Yet it is extraordinarily difficult to disentangle the two because, since the disavowal of racism during the civil rights era, research participants have become increasingly unwilling to openly express views that may be condemned as racist. Thus, the IAT could defensibly be viewed as a subtle measure of conscious psychological processes, of attitudes and beliefs that are known to oneself yet intentionally concealed from researchers. This empirical ambiguity has been practically eclipsed by the unconscious bias account. Why?

One possibility is that unconscious bias, even if not incontrovertibly demonstrated by the IAT, warrants the attention of legal scholars because it poses a unique challenge for antidiscrimination doctrine. This explanation for the ascendance of the unconscious bias discourse is intuitively appealing and widely embraced. But it is wrong. Antidiscrimination law grapples as well, or as poorly, with unconscious bias as with covert bias. Neither statutory nor constitutional antidiscrimination law turns on the distinction between the two. While the research cannot distinguish between conscious and unconscious bias, the law (fortunately) does not require courts to do so.

The better explanation for the prominence of the unconscious bias discourse relates to the comforting narrative it offers about our nation's progress in overcoming its racist history. Assertions of widespread unconscious bias are more palatable than parallel claims about covert bias. The invocation of unconscious bias levels neither accusation nor blame so much as it identifies a quasi-medical problem buried deep within us all, an ailment that distorts our thinking and behavior. People may be willing to accept that unconscious bias influences their behavior, even if they would vigorously deny harboring conscious bias. Assertions of conscious bias would open a constellation of vexing issues—for example, whether racial disparities reflect discrimination or group differences, whether discrimination may be rational, and if so whether it should be prohibited. The discussion of such matters would be uncomfortable for many and, in any event, would be unlikely to yield any quick consensus. The unconscious bias discourse promotes a (superficial) consensus that the race problem persists precisely by bypassing potential sources of disagreement.

Despite its ostensible political benefits, the unconscious bias discourse may disserve the cause of racial justice. Just as it misdescribes the IAT by eclipsing

the ambiguity of its findings, the unconscious bias approach prompts people to acknowledge the persistence of the race problem by misdescribing it. The unconscious bias approach not only discounts the persistence of knowing discrimination, it elides the substantive inequalities that fuel conscious and unconscious bias alike. While we do not doubt the existence of unconscious bias, we do doubt that contemporary racial bias accounts for all, or even most, of the racial injustice that bedevils our society. The racial injustices that most trouble us are substantive—educational failure, large-scale incarceration, segregated and impoverished communities—and stem from a complex interplay of economic, historical, political, and social influences. While historical bias has certainly played a role in producing these inequalities, it is fanciful to attribute their persistence to contemporary bias, unconscious or otherwise. The goal of racial justice efforts should be the alleviation of substantive inequalities, not the eradication of unconscious bias.

The unconscious bias discourse is as likely to subvert as to further the goal of substantive racial justice. A narrow focus on the IAT may fail as its empirical claims receive greater scrutiny, which would make it difficult for scholars who have linked their policy positions to the IAT to maintain the impartiality that is the hallmark of the scholar's commitment to truth. But even emphasizing unconscious bias more generally would be a mistake.

If people accept the eradication of unconscious bias as the goal of racial reform (and they will have to if the rhetoric is to be persuasive), they would be likely to push the theory in directions that siphon energy away from problems of substantive inequality and that may be undesirable in their own right. The unconscious bias discourse, for example, could lead either to the expanded use of diversity training, or the broad imposition of a norm of instrumentally rational decision making. Neither would further the cause of racial justice. The most fundamental problems with the unconscious bias discourse, though, are that it reinforces a misguided preoccupation with mental state, and perpetuates an obsession with antidiscrimination law, rather than policy reform, as a means of realizing racial justice goals. If the goal is to eliminate substantive inequalities, then the task of racial justice advocates should be to say why those inequalities are objectionable and how to address them. Not every claim for racial justice needs to be addressed to a court applying antidiscrimination doctrine. The best political approach, over the long term, is to straightforwardly define and defend policy goals, and then figure out how to achieve them.

* * *

Part I describes the IAT and the ambiguity of its findings. Part II explains the irrelevance of state of mind in constitutional and statutory antidiscrimination law. Part III explains the political pressures that generate the unconscious bias interpretation of the IAT. Part IV highlights the costs of the unconscious bias approach and why we reject it.

I. EMPIRICAL AMBIGUITY

This Part first provides a fuller description of the IAT and its findings. It then explains how the unconscious bias interpretation depends on rejecting the possibility of merely covert bias, which is extraordinarily difficult to do.

A. *The Measurement of Implicit Bias*

The IAT is intended to uncover “implicit bias” by measuring the strength of the association between social categories (e.g., blacks or whites) and positive and negative attributes (e.g., “joy” and “love” versus “agony” and “evil”). Although separate IATs have been developed with respect to many different traits (e.g., sex, age, nationality, weight, political affiliation, and sexual orientation), the form of the IAT that has attracted the most attention is the Race IAT,⁸ which has been taken by more than four million people.⁹ Although not the only measure of implicit bias, it is unquestionably the most widely discussed.¹⁰

⁸ Russell H. Fazio & Michael A. Olson, *Implicit Measures in Social Cognition Research: Their Meaning and Use*, 54 ANN. REV. PSYCHOL. 297, 307–08 (2003) (describing various IATs).

⁹ Cole, *supra* note 5.

¹⁰ Other researchers have used similar procedures to measure implicit bias. See Russell H. Fazio et al., *Variability in Automatic Activation as an Unobtrusive Measure of Racial Attitudes: A Bona Fide Pipeline?*, 69 J. PERSONALITY & SOC. PSYCHOL. 1013 (1995) [hereinafter Fazio et al., *Variability*]. In this study, Fazio and his colleagues developed a procedure, which they call a “bona fide pipeline,” for measuring implicit bias. *Id.* In this procedure, task participants must first classify a series of adjectives as positive or negative. Then, they study a series of faces for a subsequent memory task. Next, they see a series of previously viewed and new faces, and must indicate whether they have seen them before. Finally, after being primed with a black or white face (seeing a face presented for such a short time that it could not be recognized), they see an adjective and have to classify it as positive or negative. If a person implicitly favors white over black Americans, the test predicts that he or she will be able to classify positive adjectives faster after being primed with a white face and negative adjectives faster after being primed with a black face. See *id.*; see also Russell H. Fazio et al., *On the Automatic Activation of Attitudes*, 50 J. PERSONALITY & SOC. PSYCHOL. 229 (1986). Other researchers primed participants with the words “Black” or “White” and then showed them a string of letters that sometimes were not words but other times were words that related to positive or negative stereotypes of black and white Americans, or words that did not relate to stereotypes. Bernd Wittenbrink et al., *Evidence for Racial*

A participant completing the Race IAT first classifies a series of faces as white or black, by pressing one button on the computer keyboard for Black and another button for White.¹¹ The participant then sorts words (e.g., joy, pain, agony, happiness) into positive or negative categories (e.g., good or bad). These two sorting exercises are practice for the two trials used to compute the participant's implicit bias score.

The final two sorting trials each involve the pairing of a racial category and a positive or negative attribute. In one iteration, the negative attribute would be paired with blacks, and the positive attribute with whites.¹² A specific key would be assigned to each pairing. As words or images appear on the computer screen, the participant is instructed to press the appropriate key as quickly as possible, indicating whether the word or image belongs, for example, to the black–negative pairing or the white–positive pairing.¹³ In the next iteration, the pairings would be reversed (i.e., in our example, Black would be paired with the positive attribute and White with the negative one), and participants are again instructed to assign new words or images as quickly as possible to one pairing or the other.¹⁴ The goal is to correctly sort the words and images as quickly as possible into one pairing or the other. The test measures the time, in milliseconds, that it takes people to make these classifications by pressing the appropriate button.

The difference in response time between these two conditions is the measure of implicit bias. If a participant more quickly sorts images and words when Black is paired with the negative attribute and White with the positive attribute (compared to when the pairings are reversed), then the participant is said to have an implicit bias against African Americans. If the participant sorts words and images equally quick irrespective of which racial group is

Prejudice at the Implicit Level and Its Relationship with Questionnaire Measures, 72 J. PERSONALITY & SOC. PSYCHOL. 262 (1997). Researchers then measured the difference in time it took participants to classify stereotype-relevant and irrelevant words. *Id.*

¹¹ See Greenwald & Krieger, *supra* note 4, at 952–53; Kang, *Trojan Horses of Race*, *supra* note 4, at 1508–11; Krieger & Fiske, *supra* note 4, at 1033.

¹² In the actual Race IAT test, the formal race category labels “African American” and “European American” are used instead of “black” and “white,” as the latter labels conjure associations with “good” and “bad” independent of race. Such associations could confound measurements of bias in the test. Greenwald & Krieger, *supra* note 4, at 952 & n.24.

¹³ Participants would press one button if they saw a black face or a negative word and another button if they saw a white face or a positive word.

¹⁴ The IAT is methodologically rigorous. The order of the pairings, for example, is randomized across test administrations. For a discussion of the methodological soundness of the IAT, see Kang, *Trojan Horses of Race*, *supra* note 4, at 1528–35.

associated with which attribute, then the participant is said not to have any implicit bias.¹⁵ The idea here, which is fairly intuitive, is that someone who views white Americans more positively than black Americans will be faster on the trials that have one button for white Americans and good words and another button for black Americans and bad words, compared to trials that have one button for white Americans and bad words and another button for black Americans and bad words.

The majority of participants sort words and images faster when White is paired with the positive attribute, and Black with the negative attribute.¹⁶ The majority of participants are thus said to have an implicit bias against African Americans.¹⁷

Researchers have also administered various measures of conscious or explicit racial bias to Race IAT participants, and have found that implicit bias scores are not tightly correlated with measures of explicit bias. Some individuals even show implicit bias in the absence of any evidence of explicit bias. Researchers have attempted to extend these findings regarding the divergence in measures of implicit and explicit bias to show that implicit bias scores correlate with discriminatory behavior in cases where measures of explicit bias do not.¹⁸ Indeed, some researchers have proclaimed explicit bias

¹⁵ And finally, of course, if the participant more quickly sorts images and words when the pairings are white–negative and black–positive, then the participant is said to have an implicit bias against whites.

¹⁶ See Greenwald & Krieger, *supra* note 4, at 958 tbl.2.

¹⁷ *Id.* Note, however, that while any non-African-American racial group in the United States has a significant number of persons with an implicit bias against African Americans, African Americans themselves do not show substantial implicit bias in a particular direction. According to data gathered from the Project Implicit website, 71.5% of white participants favor European Americans, whereas only 32.4% of black participants favor European Americans. Moreover, 34.1% of black participants favor African Americans, and 33.6% show no preference for either racial group. *Id.*

¹⁸ People scoring high on implicit bias measures tend to show their bias in situations and behaviors that are less subject to conscious control or in behaviors where race is not explicitly at play in their decisions. For example, Shelton and Richeson report that white Americans' implicit biases against black Americans predicted the friendliness of their nonverbal behavior toward black partners during a game similar to tic-tac-toe. J. Nicole Shelton & Jennifer A. Richeson, *Interracial Interactions: A Relational Approach*, in *ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY* 153–54 (Mark P. Zanna ed., 2006). Dovidio, Kawakami, and Gaertner found that implicit bias predicted nonverbal friendliness of white Americans toward black-American partners during a conversation to get acquainted. John F. Dovidio et al., *Implicit and Explicit Prejudice and Interracial Interactions*, 82 *J. PERSONALITY & SOC. PSYCHOL.* 62 (2002). Similarly, another study found that implicit bias and not Modern Racism Scale (MRS) scores predicted the amount of eye contact and blinking that white participants displayed when interacting with a black partner. John F. Dovidio et al., *On the Nature of Prejudice; Automatic and Controlled Processes*, 33 *J. EXPERIMENTAL SOC. PSYCHOL.* 510 (1997); see also John B. McConahay, *Modern Racism, Ambivalence, and the Modern Racism Scale*, in *PREJUDICE, DISCRIMINATION, AND RACISM* (John F. Dovidio & Samuel L. Gaertner eds., 1986) (defining MRS scores). In

and implicit bias, as measured by the IAT, as “two . . . psychologically differentiated constructs.”¹⁹ The research thus situates implicit bias as the residue of a more overtly racist culture, one that the civil rights era transformed, cleansing our national psyche of conscious racial bias.

B. The Necessity of Measuring Conscious Bias

Although the research (and attendant controversy) has focused on the measurement of implicit bias,²⁰ the divergence between implicit and explicit bias is a crucial component of the implicit bias research. Implicit bias is only interesting to the extent that it explains something that conscious attitudes and beliefs do not. Consider the old-fashioned racist who insists, for example, that blacks and whites not be permitted to share a swimming pool. The conscious bias is easy to spot; after all, this person would declare quite forthrightly that he doesn't want blacks and whites to swim together. These views might also be informed by unconscious bias—the archetypal understandings that cause the person to view blacks as inferior, dirty, a source of contamination.²¹ Yet it would seem odd to search for the unconscious bias, or to explain the racist opposition to interracial swimming pools in terms of unconscious bias, when the conscious racism is so openly on display. Quite simply, there would be no need to probe the unconscious when the person has made his racism so plain.

What this means for psychological research is that it makes the most sense to attribute discriminatory behavior to unconscious bias only when an explanation cast in terms of conscious bias does not apply. Attributing discriminatory behavior to unconscious bias requires not only establishing a statistically significant correlation between the discriminatory behavior and

addition, McConnell and Leibold found that white participants classified by the IAT as higher in prejudice were less friendly and more uncomfortable around a black than a white experimenter. This diminished friendliness is not trivial. 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 (2001). Word, Zanna, and Cooper have shown that an employer's nonverbal behavior toward a job applicant during an interview can affect how well the applicant performs in that interview. Carl O. Word et al., *The Nonverbal Mediation of Self-Fulfilling Prophecies in Interracial Interaction*, 10 J. EXPERIMENTAL SOC. PSYCHOL. 109 (1974).

¹⁹ Mahzarin R. Banaji et al., *No Place for Nostalgia in Science: A Reply to Arkes and Tetlock*, 15 PSYCHOL. INQUIRY 279, 281 (2004).

²⁰ Hal R. Arkes & Philip E. Tetlock, *Attributions of Implicit Prejudice, or “Would Jesse Jackson ‘Fail’ the Implicit Association Test?”*, 15 PSYCHOL. INQUIRY 257 (2004); Klaus Fiedler et al., *Unresolved Problems with the “I”, the “A”, and the “T”*: A Logical and Psychometric Critique of the Implicit Association Test (IAT), 17 EUR. REV. SOC. PSYCHOL. 74 (2006); Mitchell & Tetlock, *supra* note 4, at 1023.

²¹ JOEL KOVEL, WHITE RACISM: A PSYCHOHISTORY 62, 65, 81 (1970).

unconscious bias scores. It also requires the rejection of conscious psychological processes as a cause of the discriminatory behavior. Thus, one may speak of unconscious bias as a cause of discriminatory behavior only when conscious bias is not.²² Consider a case in which a researcher finds that unconscious bias scores correlate strongly with hiring managers' discriminatory evaluation of job applicants. If the discriminatory evaluation of job applicants is predicted in precisely the same fashion by a measure of conscious bias, it would seem odd to continue to refer to the discrimination as "caused" by unconscious bias. If the discriminatory hiring managers admit that they think that African Americans are lazy, why mention unconscious bias, much less invoke it as an explanation? The hiring managers' unequivocal endorsement of such a racial stereotype would be all the explanation one would need. The attribution of discriminatory behavior to unconscious bias only makes sense when the discrimination cannot be attributed to conscious bias.²³ Unconscious bias, then, is what is left once the possibility of conscious bias has been eliminated.

To eliminate the conscious bias possibility, one must, of course, have an accurate measure of it. Herein lies a problem: it is extraordinarily difficult to accurately measure conscious racial bias if research participants are unlikely to admit to it. By definition, "racial bias" is pejorative. People do not want to admit to harboring racial bias and might, in fact, hold more racially biased views than they are willing to say. In the IAT research, levels of implicit bias consistently diverge from levels of conscious bias, but it is difficult to know whether that apparent divergence reflects a real underlying difference or is merely an artifact of the systematic understatement of levels of conscious bias. Conscious bias might well be underreported.

²² More precisely, discrimination should only be attributed to unconscious bias when variation in such discrimination across individuals is correlated with unconscious bias differently than it is correlated with conscious psychological processes.

²³ Our point here is really one about the attribution of behavior, about how we choose to assign behavior to causes. If the hiring managers who discriminate think that African Americans are lazy, to attribute that discrimination to conscious bias is not to say that the managers do not also harbor unconscious bias. Indeed, one might expect conscious and unconscious processes to operate in tandem. The conscious dislike that discriminatory hiring managers feel could well be explained based on some psychological preference for in-group members, a preference whose roots are buried in evolutionary history, far beyond any individual's conscious understanding. Nonetheless, it would be odd to choose to explain the discrimination based on unconscious processes, rather than the person's professed dislike of or negative beliefs about the group.

C. *The Social Desirability Problem*

The difficulty of eliciting conscious views regarding morally charged issues is well-known, with researchers having dubbed it the problem of “social desirability.”²⁴ Research participants may be understandably unwilling to express attitudes, beliefs, or perceptions that they know to be socially disapproved. This sort of disinclination to express consciously held attitudes and beliefs could artificially suppress levels of conscious bias, thus enlarging the apparent divergence of implicit and explicit bias.

A number of studies have shown that the degree of correlation between implicit and explicit attitude measures depends partly on the strength of social norms with respect to the group or practice at issue. For example, researchers have found a higher correlation between implicit and explicit attitudes toward being a vegetarian (a socially acceptable practice) than toward smoking (a more stigmatized practice).²⁵ Similarly, in a study of the IAT and explicit measures of attitudes, researchers found that attitudes about support for candidates in the 2000 presidential election were more highly correlated than attitudes about race or gender.²⁶ Other researchers found that participants’ implicit and explicit attitudes were more highly correlated when judging Islamic fundamentalists (whom it is currently socially acceptable to derogate) than when judging Jews (whom it is much less acceptable to denigrate).²⁷ While one would not expect explicit and implicit attitudes to be wholly uncorrelated, the fact that the strength of the correlation depends on the extent of social desirability pressures is noteworthy. It suggests that the divergence may be at least partly a consequence of the understatement of conscious views. When research participants feel free to express their negative views openly (as in the case of Islamic fundamentalists), implicit and explicit measures are more highly correlated than when there are social pressures to withhold negative sentiments (as in the case of Jews).

²⁴ ALLEN L. EDWARDS, *THE SOCIAL DESIRABILITY VARIABLE IN PERSONALITY ASSESSMENT AND RESEARCH* (1957); accord David Marlowe & Douglas P. Crowne, *Social Desirability and Response to Perceived Situational Demands*, 25 J. CONSULTING PSYCHOL. 109 (1961).

²⁵ Jane E. Swanson et al., *Using the Implicit Association Test To Investigate Attitude-Behaviour Consistency for Stigmatized Behaviour*, 15 COGNITION & EMOTION 207 (2001).

²⁶ Brian A. Nosek et al., *Harvesting Implicit Group Attitudes and Beliefs from a Demonstration Website*, 6 GROUP DYNAMICS: THEORY RES. & PRAC. 101 (2002).

²⁷ Francesca M. Franco & Anne Maass, *Intentional Control over Prejudice: When the Choice of the Measure Matters*, 29 EUR. J. SOC. PSYCHOL. 467 (1999).

Social desirability pressures unquestionably operate with respect to race-related views. Our society's moral condemnation of racism has been so thorough that few of the well-educated and middle class college students who populate research studies would voice unpopular racial views. To be branded racist carries a severe stigma,²⁸ and it is a label that everyone seeks to avoid. The pressure not to appear to be a racist could render the frequently noted divergence between Race IAT scores and measures of explicit bias partly, if not predominantly, an artifact of social desirability pressures and measurement error. This intuition is consistent with the fact that the implicit–explicit divergence is almost always a consequence of the explicit measures showing no bias, while the implicit measures do.

Moreover, one might expect social desirability pressures to be especially intense in the college laboratory settings where most social psychological studies are conducted.²⁹ Undergraduate universities tend to be liberal environments, and an important component of that liberal ethos is to be free of racial bias.³⁰ Violation of that norm might expose one to censure in the relatively closed setting of the college campus.

Studies conducted in settings where social desirability pressures are likely less intense have found higher levels of explicit bias. Consider, for example, survey studies such as the General Social Survey, where participants have little worry that they may run into the study administrator in the student union. The 1990 General Social Survey asked respondents for their opinions on the work ethic, violent behavior, and intelligence of black Americans, as compared to white Americans.³¹ Unlike with the studies conducted on college campuses, the responses spanned the full spectrum from positive to negative, and many reflected a bias against black Americans. Social desirability pressures do not necessarily account for these different results. After all, the populations are

²⁸ See RICHARD THOMPSON FORD, *THE RACE CARD: HOW BLUFFING ABOUT BIAS MAKES RELATIONS WORSE* (2008).

²⁹ David O. Sears, *College Sophomores in the Laboratory: Influences of a Narrow Data Base on Social Psychology's View of Human Nature*, 51 *J. PERSONALITY & SOC. PSYCHOL.* 515, 519 (1986).

³⁰ It would be interesting to see if the implicit–explicit divergence is smaller at relatively politically conservative universities such as Dartmouth.

³¹ Political scientists Kinder and Mendelberg, in 1995, reported on results from the 1990 General Social Survey, which was conducted using 1,372 English-speaking adults in the United States. Donald R. Kinder & Tali Mendelberg, *Cracks in American Apartheid: The Political Impact of Prejudice Among Desegregated Whites*, 57 *J. POL.* 402 (1995). Henry and Sears also found that a sample of adults in Los Angeles County endorse statements from the MRS and similar items. P.J. Henry & David O. Sears, *The Symbolic Racism 2000 Scale*, 23 *POL. PSYCHOL.* 253 (2002).

different.³² But these findings do contrast with those of many psychology studies, which tend to find uniformly low levels of conscious bias.

Additional evidence of the understatement of conscious bias as a result of social desirability pressures comes from research using the “bogus pipeline” method, which was developed more than three decades ago. Researchers found that when survey participants had electrodes attached to their skin, which ostensibly measured their true attitudes, they expressed more negative beliefs about African Americans, compared to participants without any electrodes.³³ In a more recent study, researchers had participants complete the IAT and a commonly used measure of explicit bias, but varied the information the participants received about the IAT.³⁴ Participants either (i) were told that the IAT was an accurate measure of racial prejudice, (ii) were told that it was an inaccurate measure of racial prejudice, or (iii) received no information at all about the IAT’s accuracy. The researchers found a stronger correlation of implicit and explicit measures when participants believed that the IAT would already have revealed their “true” attitudes to the experimenter than in the other two conditions. An obvious explanation for these results is that participants were more likely to express racially biased attitudes and beliefs when they thought that the IAT would get at the truth of their racial attitudes anyway, and were more likely to conceal biased sentiments when they thought they could do so successfully.³⁵

Additional evidence that people may understate their conscious racial biases comes from studies that show that the implicit–explicit divergence may be partly a consequence of individual differences in motivation to appear unprejudiced.³⁶ People who are highly motivated to control their prejudice typically report levels of explicit bias that are less correlated or even negatively correlated with their implicit bias.³⁷ In contrast, among people low in

³² The General Social Survey uses a cross-section of the American population, while nearly 70% of social psychology studies rely on college undergraduate participants, who might be more liberal and bias-free than the general population. Sears, *supra* note 29, at 519.

³³ Harold Sigall & Richard Page, *Current Stereotypes: A Little Fading, A Little Faking*, 18 J. PERSONALITY & SOC. PSYCHOL. 247, 249 (1971).

³⁴ Jason A. Nier, *How Dissociated Are Implicit and Explicit Racial Attitudes? A Bogus Pipeline Approach*, 8 GROUP PROCESSES & INTERGROUP REL. 39 (2005). This study used the MRS. *See id.*

³⁵ An alternative explanation is that the fear of being found prejudiced caused the research participants to express more prejudice. In other words, the IAT could prompt people to become aware of their own bias.

³⁶ *See* Bertram Gawronski et al., *Are “Implicit” Attitudes Unconscious*, 15 CONSCIOUSNESS & COGNITION 485 (2006).

³⁷ Bridget C. Dunton & Russell H. Fazio, *An Individual Difference Measure of Motivation to Control Prejudiced Reactions*, 23 PERSONALITY & SOC. PSYCHOL. BULL. 316, 324 (1997); Wilhelm Hofmann et al., *On*

motivation to control prejudice, implicit and explicit bias levels tend to be more closely correlated.³⁸ One study compared people's IAT scores with their explicit views on several topics and found that self-presentational concerns affected the strength of correlation between implicit and explicit attitudes.³⁹ This suggests that people for whom it is especially important to appear unbiased are especially likely to consciously suppress their bias.

In sum, the social desirability problem undermines confidence that measures of conscious bias are accurate. If actual levels of conscious bias are substantially higher than the experimental studies suggest,⁴⁰ then the correlation of implicit and conscious bias might be much greater than commonly supposed, and the utility of the notion of implicit bias much less so.

D. Recharacterizing the IAT

The great contribution of the IAT may be not that it captures a new type of bias, so much as that it employs a subtle and sophisticated means of measuring bias, which has become ever more elusive as research participants attempt to outsmart any test that would label them a racist. On this account, the IAT would be the most recent entry in a long line of subtle measures of racial bias.⁴¹ During the past seventy or so years, social scientists have used ever more restrained measures of racial bias in response to the methodological challenge posed by social desirability pressures. As research participants became increasingly less willing to express their (racist) views openly and honestly, straightforward measures of racial attitudes gave way to more subtle measures, in the hope of getting at bias that research participants would prefer to conceal.

Implicit-Explicit Consistency: The Moderating Role of Individual Differences in Awareness and Adjustment, 19 EUR. J. PERSONALITY 25, 45 (2005).

³⁸ Dunton & Fazio, *supra* note 37, at 324; Hofmann et al., *supra* note 37, at 45.

³⁹ Brian A. Nosek, *Moderators of the Relationship Between Implicit and Explicit Evaluation*, 134 J. EXPERIMENTAL PSYCHOL.: GEN. 565, 566 (2005).

⁴⁰ This assertion derives support from studies conducted with noncollege students in a more anonymous setting outside the psychological laboratory, from the "bogus pipeline" findings that people will admit to their implicit attitudes if they think the experimenter already knows them.

⁴¹ One might describe the bias measured by the IAT as conscious, or one might simply describe it as racial bias, without according any weight to the conscious/unconscious distinction. Alternatively, one might depict the IAT as a measure of automatic or spontaneous bias, which need not imply whether it is beyond individual awareness.

When researchers began to study racial attitudes in the late 1920s,⁴² their approach was straightforward: they asked people what they thought about different groups, and people told them.⁴³ In a landmark study, Katz and Braly gave one hundred Princeton students a list of adjectives (e.g., scientifically minded, artistic, superstitious, pugnacious, sportsmanlike, shrewd, intelligent, cruel, and physically dirty) and a list of social groups that included Germans, Italians, “Negroes,” Irish, English, Jews, Americans, Chinese, Japanese, and Turks.⁴⁴ When they asked the students to indicate which adjectives best described the groups, the students did so, characterizing some groups with negative traits. In a subsequent study, Princeton undergraduates reported differences in their willingness to have contact with groups.⁴⁵ Even after the Holocaust led psychologists to view racial stereotypes and dislikes more negatively,⁴⁶ many people remained willing to explicitly admit to negative attitudes toward other groups.

⁴² Samelson argues that as immigration to the United States declined in the mid- to late-1920s and the Nazis rose to power in Europe, psychologists shifted their focus from biological group differences to the study of attitudes toward other groups. This change from the study of biological differences to the study of attitudes occurred gradually. For example, in Carl Murchison’s 1935 edition of *A Handbook of Social Psychology*, the predecessor to the current *Handbook of Social Psychology* series, stereotypes receive only passing mention in a chapter on “Attitudes.” In contrast, there are several chapters on the social behavior of different racial groups. Franz Samelson, *From “Race Psychology” to “Studies in Prejudice”*; *Some Observations on the Thematic Reversal in Social Psychology*, 14 J. HIST. BEHAV. SCI. 265 (1978); see Gordon W. Allport, *Attitudes*, in *A HANDBOOK OF SOCIAL PSYCHOLOGY* 798 (Carl Murchison ed., 1935); Daniel Katz & Kenneth W. Braly, *Racial Prejudice and Racial Stereotypes*, 30 J. ABNORMAL & SOC. PSYCHOL. 175 (1935) [hereinafter Katz & Braly, *Racial Prejudice*]; Daniel Katz & Kenneth W. Braly, *Racial Stereotypes of One Hundred College Students*, 28 J. ABNORMAL & SOC. PSYCHOL. 280 (1933) [hereinafter Katz & Braly, *One Hundred College Students*].

⁴³ Bogardus conducted some of the first studies measuring people’s attitudes toward other races. In the first study, participants categorized racial and national groups based on the degree to which they felt friendly, neutral, or aversive toward them. Emory S. Bogardus, *Social Distance and Its Origins*, 9 J. APPLIED SOC. 216 (1925) [hereinafter Bogardus, *Origins*]. In a second study, participants reported on how socially close or distant they wished to be from the same groups, indicating whether they would allow individuals from that group to marry into their family, join their personal clubs, move to their street, take a job at their place of employment, become a citizen of their country, visit their country, or not enter their country at all. Emory S. Bogardus, *Measuring Social Distances*, 9 J. APPLIED SOC. 299 (1925). In both cases, “Turks,” “Negroes,” “Mulattoes,” and “Hindus” were at the bottom of the list, while “Canadians” and “English” were at the top. See *id.*; Bogardus, *Origins*, *supra*.

⁴⁴ Katz & Braly, *One Hundred College Students*, *supra* note 42.

⁴⁵ Katz & Braly, *Racial Prejudice*, *supra* note 42.

⁴⁶ In 1954, Gordon Allport published *The Nature of Prejudice*, a comprehensive (and now classic) review of psychological research on racial and ethnic discrimination. *The Nature of Prejudice* outlines an extensive list of sources of prejudice, ranging from the desire to scapegoat to social structure, and from personality factors to aggression and frustration. Allport sought both to define the field of the psychology of prejudice, and to use the research to reduce prejudice. See GORDON W. ALLPORT, *THE NATURE OF PREJUDICE* (1954). In 1950, Theodore Adorno and colleagues developed the idea of an authoritarian personality, characterized by

Only after the civil rights struggles of the 1950s and 1960s did the moral condemnation of racial animus spread to the general population. By the 1970s, researchers recognized that social pressures might cause research participants to express only views that they perceived as socially desirable. Psychologists sought methods of gauging prejudice that were more subtle than asking direct questions. One subtle measure that became popular was the Modern Racism Scale (MRS), which was thought to be less susceptible to social norms.⁴⁷ The MRS asks research participants to rate their agreement with a variety of statements—relating to crime, civil rights activism, and judicially mandated desegregation—that, at the time, did not directly and obviously implicate the prevailing moral condemnation of racism.⁴⁸ The MRS remains a widely used paper-and-pencil measure of racial bias,⁴⁹ but other scales also exist. The Pro-

acceptance of social conventions, identification with and submission to authority, and high levels of prejudice. See THEODORE W. ADORNO ET AL., *THE AUTHORITARIAN PERSONALITY* (1950).

⁴⁷ John B. McConahay, *Modern Racism, Ambivalence, and the Modern Racism Scale*, in PREJUDICE, DISCRIMINATION, AND RACISM (John F. Dovidio & Samuel L. Gaertner eds., 1986); John B. McConahay et al., *Has Racism Declined in America? It Depends on Who Is Asking and What Is Asked*, 25 J. CONFLICT RESOL. 563, 568 (1981) [hereinafter McConahay et al., *Has Racism Declined in America?*]. McConahay contrasted his approach with Old Fashioned Racism, which included the following sorts of statements:

I favor strong open housing laws that permit minority persons to rent or purchase housing even when the owner does not wish to rent or sell.
 It is a bad idea for blacks and whites to marry one another.
 It was wrong for the United States Supreme Court to outlaw segregation in its 1954 decision.
 If a black family with about the same income and education as I have moved next door, I would mind it a great deal.
 Generally speaking, I favor full racial integration.
 Black people are generally not as smart as whites.

Id. (responses omitted).

⁴⁸ The MRS includes the following statements:

It is easy to understand the anger of black people in America.
 Blacks have more influence upon school desegregation plans than they ought to have.
 The streets are not safe these days without a policeman around.
 Blacks are getting too demanding in their push for equal rights.
 Over the past few years blacks have gotten more economically than they deserve.
 Over the past few years the government and news media have shown more respect to blacks than they deserve.

McConahay et al., *Has Racism Declined in America?*, *supra* note 47, at 568 (responses omitted).

⁴⁹ There is some evidence that explicit questionnaire methods can predict biased behavior. Dovidio, Kawakama, Johnson, Johnson, and Howard found that MRS scores predicted white participants' reports about their experience with a black experimenter and their decisions as members of a mock jury. John F. Dovidio et al., *On the Nature of Prejudice: Automatic and Controlled Processes*, 33 J. EXPERIMENTAL SOC. PSYCHOL. 510, 517–18 (1997). Additionally, Fazio, Jackson, Dunton, and Williams found that MRS scores predicted white Americans' views about the legitimacy of the Rodney King verdict and the black community's anger in response to this verdict. Fazio et al., *Variability*, *supra* note 10, at 1018–19; see also Donald R. Kinder &

Black and Anti-Black scales,⁵⁰ for example, attempt to measure sympathy (or lack thereof) for black Americans by asking participants to rate their agreement with various quasi-factual statements about, for example, whether blacks experience discrimination or should be given special consideration in hiring.⁵¹

Although these sorts of survey questions were more subtle than the questions used in earlier racial attitudes research, researchers soon became concerned that the newer approaches were also tainted by social desirability pressures. Just as it had become impermissible to declare blacks genetically inferior to whites, so too did it later become unfashionable to criticize black culture, to say that blacks did not work as hard as whites, or to question the existence of discrimination.

In recent years, social psychologists have developed a number of less obtrusive or indirect measures of racial bias,⁵² including facial electromyography—the measurement of facial muscle movements to determine whether one liked a conversation partner.⁵³ Other researchers examined activity in the amygdala (an area of the brain associated with emotion and fear),⁵⁴ and cardiovascular reactions.⁵⁵ The characterization of racial bias as conscious or unconscious was not a prominent feature of these studies.⁵⁶ Their goal was not to distinguish unconscious from conscious bias

David O. Sears, *Prejudice and Politics: Symbolic Racism Versus Racial Threats to the Good Life*, 40 J. PERSONALITY & SOC. PSYCHOL. 414 (1981); James B. Kleugel & Eliot R. Smith, *Beliefs About Stratification*, 7 ANN. REV. SOC. 29 (1981); Donald O. Sears, *Symbolic Racism*, in ELIMINATING RACISM: PROFILES IN CONTROVERSY 53 (P.A. Taylor & D.A. Katz eds., 1988).

⁵⁰ Irwin Katz & R. Glen Hass, *Racial Ambivalence and American Value Conflict: Correlational and Priming Studies of Dual Cognitive Structures*, 55 J. PERSONALITY & SOC. PSYCHOL. 893 (1988).

⁵¹ *Id.*

⁵² See Fazio & Olson, *supra* note 8, at 297 (reviewing a range of methods for measuring implicit bias).

⁵³ Eric J. Vanman et al., *The Modern Face of Prejudice and Structural Features that Moderate the Effect of Cooperation on Affect*, 73 J. PERSONALITY & SOC. PSYCHOL. 941 (1997).

⁵⁴ See Allen J. Hart et al., *Differential Response in the Human Amygdala to Racial Outgroup vs Ingroup Face Stimuli*, 11 NEUROREPORT 2351 (2000); Elizabeth A. Phelps et al., *Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation*, 12 J. COGNITIVE NEUROSCI. 729, 731 (2000) (reporting researchers' findings that among white research participants, the amygdala was more active when the participant was looking at a black face than at a white face).

⁵⁵ See Wendy Berry Mendes et al., *Challenge and Threat During Social Interactions with White and Black Men*, 28 PERSONALITY & SOC. PSYCHOL. BULL. 939 (2002) (reporting researchers' finding that white participants who interacted with a black partner experienced cardiovascular reactions that normally appear during threatening situations).

⁵⁶ The study by Phelps and colleagues did find that participants with greater anti-black bias as measured by the IAT (but not as measured by other paper-and-pencil tests of bias) exhibited more activation in an area of the brain associated with fear and emotion when they viewed a black face, compared to when they viewed a white face. Phelps et al., *supra* note 54, at 730–31.

so much as to sidestep the impediment of social desirability pressures. Even as the measures became ever more subtle, they were assumed to detect the same bias that earlier generations had expressed forthrightly.

Even if the IAT does not offer incontrovertible evidence of unconscious bias, one might still justify the attention accorded unconscious bias by asserting that it poses a unique challenge for antidiscrimination law. That assertion, however, would be wrong.

II. ANTIDISCRIMINATION DOCTRINE

The theory of unconscious bias is compelling. Narratives of the unconscious are always revelatory: Freudian dream analysis reveals the hidden causes of human phobias and neuroses; the erratic behavior of Hitchcock's otherwise virtuous Marnie is deciphered when the roots of her kleptomania are revealed in an abusive childhood; the IAT reveals the true source of racial injustice in a post-civil rights world. The value of such revelation is to serve prescription—indeed it is to determine the appropriate prescription. To take another example, once we know that the analysand's hysteria is rooted in the Electra complex, we can begin to cure it by helping her to come to a healthier relationship with her father and by extension, men in general; once we understand the source of Marnie's compulsion, we are confident a healthy psychology cannot be far away. Faced with such a powerful revelation, it's only natural to want to do something with it; to use this information to fix some mistake that we've made in our ignorance. But what if knowledge of the unconscious can't help us cure the patient?

In this Part, we show that in statutory and constitutional antidiscrimination doctrine, nothing much turns on the question of whether bias is conscious. While experimental research cannot distinguish between unconscious and covert bias, the legal doctrine does not require courts to do so. In constitutional and statutory law alike, the courts confront the difficult problem of discrimination that is not overt. The challenge posed by covert discrimination is substantial, but the nature and magnitude of that challenge does not depend on whether such discrimination is prompted by implicit rather than conscious racial attitudes or beliefs.

A. *Statutory Law: Title VII*

The concept of unconscious bias is largely irrelevant to any question of unlawful discrimination under Title VII, the federal statute that prohibits discrimination in the workplace.⁵⁷ Under Title VII, the courts have developed a doctrinal framework that does not require the fact finder to make any determination as to the state of mind of the defendant. The prospect of unconscious bias might matter if (i) the law considered any single legitimate conscious motivation sufficient to definitively settle the case in favor of the defendant or (ii) the law could not comprehend a defendant that was unaware of its own motivations. But neither of these is the case. The law has developed a method for dealing with cases of mixed motives: there is no reason one or more of the motives in the mix could not be unconsciously held, provided it influenced the challenged decision. Moreover, the law applies to multi-member decision-making bodies, where partial unawareness of motivations is quite common.

1. *Single-Motive Cases: Formalism, Realism, and Employment at Will*

Cases that involve overtly discriminatory policies or direct statements evidencing discrimination by a decision maker have declined in frequency during the past few decades. Most claims of discrimination concern concealed bias, where the plaintiff must prove discrimination by inference.⁵⁸ The simplest such claim is one in which racial bias is alleged as the only motivation for the challenged action. There is a well-developed formal structure for analyzing such claims under Title VII. Under this structure, it makes no difference whether the concealed bias is unconscious or merely covert.

Proof in single-motive cases works by a deductive process: if other plausible motivations for the challenged action are disproved, the fact finder may infer that the motivation was unlawfully discriminatory. According to the U.S. Supreme Court, evidence of unlawful intentional discrimination in Title VII litigation shall unfold according to a formal structure: Plaintiff alleges (to survive 12(b)(6)) or proves (to avoid judgment for defendant as a matter of law) his prima facie case. Under the Court's opinion in *McDonnell Douglas*

⁵⁷ 42 U.S.C. § 2000e (2000).

⁵⁸ See, e.g., Sean W. Colligan, *In Good Measure: Workforce Demographics and Statistical Proof of Discrimination*, 23 LAB. LAW. 59, 59 (2007) ("Courts have long recognized that plaintiffs in discrimination lawsuits rarely have the benefit of direct evidence of discrimination. Because overt expressions of discriminatory motives by managers are rare, indirect evidence of discriminatory intent is often crucial in employment cases." (footnote omitted)).

Corp. v. Green,⁵⁹ in order to establish a prima facie case, the plaintiff must show:

- (i) that he belongs to a racial minority;
- (ii) that he applied and was qualified for a job for which the employer was seeking applicants;
- (iii) that, despite his qualifications, he was rejected; and
- (iv) that, after his rejection, the position remained open and the employer continued to seek applicants from persons of complainant's qualifications.⁶⁰

The defendant can either challenge an element of the prima facie case or defend the challenged action as motivated by a legitimate, nondiscriminatory reason. If the defendant does neither, the plaintiff is entitled to judgment as a matter of law. If the defendant defends the challenged action as motivated by a nondiscriminatory reason, plaintiff has the opportunity to rejoin that the proffered reason is a pretext.⁶¹

Notice that the prima facie case of intentional discrimination does not require the plaintiff to prove intentional discrimination. And, in theory, if the defendant does nothing after the plaintiff establishes his prima facie case, the plaintiff is entitled to judgment as a matter of law.⁶² Of course, it is unlikely that an employer would not respond to the plaintiff's prima facie case.⁶³ The threat of judgment as a result of an unanswered prima facie case is a bluff, designed to force the defendant to answer rather than demur. The *McDonnell Douglas* framework acknowledges that discrimination will rarely be proven directly, but rather will most often have to be inferred from circumstances. Were defendants able to simply demur, they would almost always prevail. The second "stage" of the framework is designed to force defendants to explain suspicious circumstances, in the hope that hidden discriminatory motives will be revealed or suggested by unconvincing alibis. The final potential "stage" in

⁵⁹ 411 U.S. 792, 802 (1973).

⁶⁰ *Id.*

⁶¹ *Id.* at 804.

⁶² This would seem to contradict either the well-worn observation that the plaintiff in a civil action carries the burden of persuasion or the understanding of the cause of action as one alleging intentional discrimination.

⁶³ In fact, there is not a single reported case in which a defendant did nothing in response to a plaintiff's prima facie case. Indeed it seems that no court imagined that any defendant would do nothing. Judge Denny Chin & Jodi Golinsky, *Moving Beyond McDonnell Douglas: A Simplified Method for Assessing Evidence in Discrimination Cases*, 64 BROOK. L. REV. 659, 665 (1998).

the production of evidence occurs after the defendant has offered a nondiscriminatory reason for the challenged action. The plaintiff can rejoin that the proffered reason is a “pretext.”

So, in theory the plaintiff will win if he makes out a prima facie case of discrimination, and if the defendant fails to produce evidence that it took the adverse action for a legitimate, nondiscriminatory reason. At this point, the plaintiff has not offered any evidence of intentional discrimination (other than the prima facie case)—he certainly hasn’t “proven” discrimination by any conventional standard of proof. It’s plausible that no reasonable fact finder would conclude from the evidence presented that the defendant discriminated. Although supposedly the plaintiff bears the burden of proof, he can still win without coming close to proving the ultimate factual question if the defendant fails to allege a nondiscriminatory reason for the challenged action.

On this reading, “discriminatory intent” is simply shorthand for the absence of a good reason for the challenged action. Indeed, this is a sensible way of interpreting the framework that the Supreme Court put forth in *McDonnell Douglas* and further clarified in *Texas Department of Community Affairs v. Burdine*⁶⁴: if the plaintiff convinces the fact finder that the legitimate nondiscriminatory reason offered by the defendant–employer was not the real reason for the adverse action, the plaintiff is entitled to judgment as a matter of law.⁶⁵ In other words, the fact finder was required to infer discriminatory intent from the prima facie case and the failure of the defendant’s proffered nondiscriminatory reason.⁶⁶ Through this deductive method, intent is not a thing to be discovered, but rather is revealed by an absence—the lack of any other credible reason for the adverse employment action leaves intentional discrimination as the only acceptable inference.

Such an interpretation of the doctrine avoids the question of discriminatory motivation altogether. If the question of motivation plays no role in the inquiry, then the prospect of unconscious bias cannot play a role either. Let’s call such an approach to the doctrine “formalist.” The formalist approach

⁶⁴ Tex. Dep’t of Cmty. Affairs v. Burdine, 450 U.S. 248, 252–58 (1981).

⁶⁵ Justice Blackmun articulated this approach in his concurrence in *U.S. Postal Service Board of Governors v. Aikens*, 460 U.S. 711, 717–18 (1983) (Blackmun, J., concurring); see also Catherine J. Lanctot, *The Defendant Lies and the Plaintiff Loses: The Fallacy of the “Pretext-Plus” Rule in Employment Discrimination Cases*, 43 HASTINGS L.J. 59 (1991).

⁶⁶ See, e.g., *Griffin v. George B. Buck Consulting Actuaries, Inc.*, 551 F. Supp. 1385, 1389–91 (S.D.N.Y. 1982) (finding for the plaintiff on the basis of the failure of the defendant’s proffered nondiscriminatory reason).

recognizes that, practically speaking, there may often be no way to prove discrimination other than through some deductive process with a formal structure of inferences. Moreover, a positive account of what discrimination “is” comes with all sorts of conceptual difficulties. The formalist approach of *McDonnell Douglas* happily avoids confronting the inevitable ideological conflict over the appropriate meaning of discrimination by sidestepping the question altogether—we simply presume that unjustified negative actions directed at members of historically despised or excluded groups are “discriminatory.”

The mandatory inference of the *McDonnell Douglas* approach has been harshly criticized. Some commentators have complained that the prima facie case is not rigorous enough to justify judgment for the plaintiff and risks imposing liability, not only for discrimination, but for arbitrariness, mistakes, and bad judgment.⁶⁷ This critique presupposes some positive account of discrimination. From this epistemologically “empiricist” position, discrimination is an empirical fact in the world, waiting to be discovered.

In *St. Mary’s Honor Center v. Hicks*,⁶⁸ the Supreme Court departed from the formalism of the *McDonnell Douglas* line of cases and embraced a more “empiricist” position. However, underlying the move to epistemological empiricism was a policy objective: protecting the employer’s prerogative to terminate the employment relationship at will and avoiding a slide into a de facto good cause obligation.

Melvin Hicks was fired from his job as a shift commander at Saint Mary’s Honor Center—a halfway house for parolees in Missouri—after what looked like a long vendetta against him led to a demotion and culminated in a heated exchange with his immediate supervisor.⁶⁹ Hicks sued Saint Mary’s for racial discrimination. Saint Mary’s responded that Hicks had been demoted because he didn’t effectively discipline the employees under his direct supervision and that he had been fired because he threatened his immediate supervisor. The district court didn’t buy this excuse: Hicks was the only supervisor disciplined for rule violations by subordinates, even though such violations were common. And the supervisor whom Hicks had supposedly threatened had picked a fight

⁶⁷ See, e.g., Deborah C. Malamud, *The Last Minuet: Disparate Treatment After Hicks*, 93 MICH. L. REV. 2229 (1995).

⁶⁸ 509 U.S. 502 (1993).

⁶⁹ *Hicks v. St. Mary’s Honor Ctr.*, 756 F. Supp. 1244, 1246–49 (E.D. Mo. 1991), *rev’d*, 970 F.2d 487, 492 (8th Cir. 1992), *rev’d*, 509 U.S. 502 (1993).

with Hicks in order to provoke him. In the court's view, Saint Mary's had not offered a convincing explanation for demoting or firing Hicks.⁷⁰ But the district court didn't think Hicks had proven that he had been fired because of his race: "Although [Hicks] has proven the existence of a crusade to terminate him, he has not proven that the crusade was racially rather than personally motivated."⁷¹ The Eighth Circuit Court of Appeals reversed. Applying the *McDonnell Douglas* method, it reasoned that, "because all of defendants' proffered reasons were discredited, defendants were in a position of having offered . . . no rebuttal to an established inference that they had discriminated against plaintiff on the basis of his race."⁷²

In reversing the court of appeals, Justice Scalia, writing for a five-Justice majority, insisted that the plaintiff must convince the fact finder that the defendant, in fact, discriminated in order to prevail.⁷³ After *Hicks*, the fact finder was not required to infer intent from the prima facie case and the falsity of the proffered nondiscriminatory reason. Despite the failure of the proffered nondiscriminatory reason, the Court in *Hicks* allowed that the fact finder might be unpersuaded as to discriminatory intent. Justice Scalia suggested that some nondiscriminatory reason that the defendant did not "allege"—perhaps because its agents were embarrassed to admit it—might have in fact motivated the adverse action.⁷⁴ One can't deduce from the failure of the proffered explanation that the defendant was motivated by discriminatory intent: lots of other reasons—not alleged at trial but also not ruled out—may have motivated the challenged action.

The Court's opinion in *Hicks* situated discrimination as an empirical fact to be discovered, something that a fact finder could discover or that a plaintiff could prove, rather than an artifact of a formal structure of deduction, or the product of a refined juridical narrative. This epistemologically empiricist position is a plausible interpretation of the doctrine. The plaintiff bears the burden of persuasion: whatever the evidence, the plaintiff must convince the fact finder that the defendant discriminated. The failure of the defendant's proffered nondiscriminatory reasons doesn't necessarily prove the fact of discrimination. At most, it suggests that some of the agents of the defendant

⁷⁰ *Id.* at 1250–51.

⁷¹ *Id.* at 1252.

⁷² *Hicks v. St. Mary's Honor Ctr.*, 970 F.2d 487, 492 (8th Cir. 1992), *rev'd*, 509 U.S. 502 (1993).

⁷³ *St. Mary's Honor Ctr. v. Hicks*, 509 U.S. 502, 511, 514–15 (1993).

⁷⁴ *Id.* at 513–14.

(which, as Justice Scalia points out,⁷⁵ is almost always not itself a person but rather a corporate entity) perjured themselves on the stand, in depositions or in affidavits; it may only suggest that the agents were misinformed or that other documents were misleading or simply that the evidence—maybe true, maybe not—wasn't believed. The fact finder could disbelieve the defendant's evidence yet still doubt that discrimination occurred. Perhaps other evidence—even evidence that the plaintiff thought bolstered his case—persuades him that some third, unarticulated nondiscriminatory reason motivated the adverse action; or perhaps because when all is said and done, the plaintiff's story of discrimination just doesn't ring true. If the plaintiff hasn't proven his case, he isn't entitled to relief, regardless of what the defendant has or hasn't proven.

Justice Souter, in his dissent in *Hicks*, adopts a more epistemologically formal position, emphasizing that a finding of intentional discrimination must be seen as the outcome of the formal process:

[P]roof of a prima facie case not only raises an inference of discrimination; in the absence of further evidence, it also creates a mandatory presumption in favor of the plaintiff. . . . [O]nce the plaintiff [proves his prima facie case] the employer must either respond or lose. . . . Thus, if the employer remains silent because it acted for a reason it is too embarrassed to reveal, or for a reason it fails to discover, the plaintiff is entitled to judgment⁷⁶

Here again, we see a conception of discrimination as an artifact of a formal process: if the plaintiff has proven his prima facie case and no credible nondiscriminatory reason is offered, then, *ipso facto*, the adverse employment action was “discriminatory.”

Justice Souter complained that under Justice Scalia's majority opinion, “once a Title VII plaintiff succeeds in showing . . . that the defendant has come forward with pretextual reasons for its actions . . . , the factfinder still may proceed to roam the record, searching for some nondiscriminatory explanation that the defendant has not raised”⁷⁷ For him, the majority's rejection of a mandatory inference of discrimination placed an unfair burden on plaintiffs. It seemed to “saddle the victims of discrimination with the burden of either

⁷⁵ *Id.* at 520.

⁷⁶ *Id.* at 528 (Souter, J., dissenting) (citations omitted).

⁷⁷ *Id.* at 525.

producing direct evidence of discriminatory intent or eliminating the entire universe of possible nondiscriminatory reasons.”⁷⁸

From Justice Scalia’s empiricist perspective, this is the only sensible way to proceed: of course the fact finder should be able to roam the record—in other words, consider all of the evidence—and may, indeed must, determine based on that evidence whether it is persuaded that the employer intentionally discriminated. And from this perspective, the use of the term “pretextual” is question begging: the proffered reasons are not “pretextual” under *McDonnell Douglas* and *Burdine* simply because they are not believed; they are pretextual only if they are a pretext *for* discriminatory intent.

The disagreement between Justices Scalia and Souter may appear to involve only technical procedural and evidentiary questions: What must a fact finder infer from the failure of the defendant’s alibi? Is it “‘more likely than not’ that the employer who lies [or is disbelieved] is simply trying to cover up the illegality alleged by the plaintiff”?⁷⁹ But the central dispute is conceptual and ideological. It centers on the question of the status of at-will employment. Justice Scalia’s opinion in *Hicks* reflected characteristic conservative concerns. It was predictable that some supervisors would not confess to a host of nondiscriminatory but ignoble motives and instead offer unconvincing excuses. And even if they did confess, many juries, outraged at such unfair—but not discriminatory—treatment, would find the defendants liable anyway. A rule that automatically made the employer liable when a jury or judge didn’t buy its explanation might capture more discrimination, but it would do so only by turning Title VII into a virtual guarantee of life tenure for members of protected groups. In practice, such a rule would substitute a for-cause termination requirement for the narrow prohibition of discriminatory firing. In Justice Scalia’s view, this was too high a price to pay: better to let some bad actors get away with it.

Justice Souter’s dissent in *Hicks* reflected a lawyer’s sense of the practical limitations of fact-finding. Discriminatory intent is extremely hard to prove. Occasionally an employer slips up and announces his prejudice in mixed company or memorializes it in a memo. But this is rare—when it does happen, it makes the newspapers. The *McDonnell Douglas* method offered a sensible, fair, and predictable way to uncover discrimination that employers had every incentive to hide. For Justice Souter, Justice Scalia’s opinion in *Hicks*

⁷⁸ *Id.* at 528.

⁷⁹ *Id.* at 536 (quoting *Furnco Constr. Corp. v. Waters*, 438 U.S. 567, 577 (1978)).

undermined an orderly procedure and replaced it with a free-for-all. The employer had a chance to explain suspicious behavior and an obligation to present accurate testimony at trial. If it failed to do either, it was reasonable to assume the worst. And if the employer acted for bad but nondiscriminatory reasons, they are still bad reasons: let the employer fess up and face the jury or keep quiet and accept liability.

What is important to note is that nothing turns on the existence of unconscious bias in either the formalist or empiricist account. In the formalist account, the question of unconscious bias is irrelevant because, to put it starkly, state of mind simply does not matter. To decide whether the defendant discriminated does not require the fact finder to probe anyone's mental processes, so much as to decide whether the proffered legitimate, nondiscriminatory reason is credible. If it is not, the defendant loses—simple as that. In the empiricist account, the fact finder must be convinced of the existence of a discriminatory action, but nothing in the doctrine directs the fact finder to attend to the difference between conscious and unconscious motivation. Put differently, the doctrine draws no distinction between discriminatory purpose (where the discriminatory aim is conscious) and discriminatory motive (where it is plausibly not). The single-motive structure, then, cannot be criticized for excluding discrimination that results from unconscious bias.

Justice Souter's worry in *Hicks* stems from a profound, and entirely justified, lack of faith in the ability of plaintiffs to prove discriminatory intent. Without "direct evidence," the only way a plaintiff can prove discriminatory intent is by process of elimination—a process made manageable by the formal structure and mandatory inference undermined by *Hicks*—the plaintiff doesn't have to eliminate every conceivable nondiscriminatory reason for the challenged action—only the reason the defendant put forward. Justice Souter saw that the real implication of Justice Scalia's epistemological realism was to press toward a universal direct evidence requirement: indeed, after *Hicks* and until the Court clarified its position in *Reeves v. Sanderson Plumbing Products*,⁸⁰ some appeals courts followed a "pretext-plus" rule that required plaintiffs not only to prove that the defendant's proffered rationale was a pretext, but also to offer some additional evidence of discriminatory intent

⁸⁰ 530 U.S. 133 (2000).

before the fact finder would be entitled to infer unlawful discrimination.⁸¹ The problems with pretext-plus—indeed the problems with any direct evidence requirement—go beyond the practical concern that direct evidence is difficult to obtain. In many cases, direct evidence is conceptually elusive because it is unclear what such evidence would consist of.⁸²

The facts of *Hicks* demonstrate this. The court of appeals in *Hicks* found that although Melvin Hicks had demonstrated that the reasons the defendant offered for demoting and later firing him were pretextual, the real reason was not Hicks's race but rather a personal grudge against him. Of course, it's true that a personal grudge could be entirely unrelated to race. But at the same time, it could be intimately tied up with race. At one point, because of the escalating tensions between Hicks and his supervisors, a supervisor accused Hicks of threatening him.⁸³ Did the supervisor find Hicks's behavior threatening, in part, because of common stereotypes about aggressive and violent black men? The court of appeals in *Hicks* thought of a personal grudge against Hicks as an alternative explanation to racial discrimination, but one could easily see it as evidence of racial discrimination.

2. *Mixed-Motive Cases: Causation and Duty*

The Supreme Court first articulated the mixed-motive framework in *Price Waterhouse v. Hopkins*.⁸⁴ Ann Hopkins filed suit against the accounting firm Price Waterhouse, alleging sex discrimination in the firm's failure to promote her to partner. Price Waterhouse claimed Hopkins was passed over for partnership due to poor interpersonal skills; Hopkins argued that the real reason she was passed over was her sex. Hopkins cited a number of comments that made note of her sex and slipped into sex stereotypes. One partner said she was too "macho."⁸⁵ Another advised her to go to "charm school."⁸⁶ The partner who explained the company's decision to delay consideration of her for promotion offered the kind of advice a beauty pageant hopeful might expect:

⁸¹ See MICHAEL ZIMMER ET AL., *CASES AND MATERIALS ON EMPLOYMENT DISCRIMINATION* 146 (6th ed. 2003).

⁸² See generally Charles A. Sullivan, *Accounting for Price Waterhouse: Proving Disparate Treatment Under Title VII*, 56 BROOK. L. REV. 1107, 1138 (1991) (arguing that a "direct evidence" requirement is incoherent).

⁸³ *Hicks v. St. Mary's Honor Ctr.*, 970 F.2d 487, 489 (8th Cir. 1992), *rev'd*, 509 U.S. 502 (1993).

⁸⁴ 490 U.S. 228 (1989).

⁸⁵ *Id.* at 235.

⁸⁶ *Id.*

he told Hopkins to “walk more femininely, talk more femininely, dress more femininely, wear make-up, have her hair styled, and wear jewelry.”⁸⁷

The single-motive approach seems inadequate in a case like *Price Waterhouse* because, even as there was clear evidence of sexism, it was also the case that the firm really was concerned about Hopkins’s interpersonal skills and her ability to get along well with co-workers.⁸⁸ These concerns were not a pretext for discrimination, hastily concocted after the fact, in preparation for trial. These concerns were unquestionably legitimate. At the same time, however, the legitimacy of these concerns should not be allowed to obscure the sexism that also played a role in the decision. It’s impossible to imagine that the firm would have counseled a man that his partnership chances could be improved by a course at charm school. The mixed-motive structure accommodates the possible coexistence of legitimate and sexist considerations in the decision-making process. The mixed-motive approach does not require the fact finder to reject the employer’s proffered legitimate, nondiscriminatory reason in order to find for the plaintiff. The fact finder might believe that both the employer’s stated reason and a discriminatory motive played a role in the challenged adverse action.

The issue that divided the Justices in *Price Waterhouse* concerned how substantial a role the discriminatory consideration must play in the challenged action in order for liability to attach. All nine Justices agreed that Hopkins had to show she was passed over for partner because of her sex.⁸⁹ But the Court fractured over what “because of” meant. The dissenting Justices insisted that “because of” meant that the prohibited motive was the “but for” cause of the challenged decision. On this view, Hopkins needed to prove that, were it not for sex discrimination, she would have been promoted to partner.⁹⁰

⁸⁷ *Id.*

⁸⁸ Of course, as in *Hicks*, the firm’s legitimate concerns about Hopkins’s interpersonal skills may well have been intertwined with sexism. Sexist attitudes may have been a big part of why she had trouble getting along with others.

⁸⁹ Title VII forbids an employer from making certain employment-related decisions “because of” the “race, color, religion, sex, or national origin” of an employee or applicant. 42 U.S.C. §§ 2000e-2(a)(1), (2) (2006).

⁹⁰ *See Hopkins*, 490 U.S. at 281–82, 284 (Kennedy, J., dissenting). These Justices thought that, on the facts, Hopkins had only shown that discrimination was “in the air”—some partners had expressed sexist views that may or may not have affected her candidacy. *See id.* at 293–95. *But see id.* at 241, 251 (plurality opinion) (rejecting *Price Waterhouse*’s “in the air” formulation and arguing that discrimination was “brought to ground and visited upon” Hopkins).

In a plurality opinion, Justice Brennan rejected such a requirement of “but for” causation, offering a physical analogy to explain why he did:

Suppose two physical forces act upon and move an object, and suppose that either force acting alone would have moved the object. As the dissent would have it, *neither* physical force was a “cause” of the motion unless we can show that but for one or both of them, the object would not have moved Events that are causally overdetermined . . . may not have any “cause” at all. This cannot be so.⁹¹

He concluded that if sexism was a factor in the decision, it was reasonable to say that sexism “caused” it. For Justice Brennan, a decision is “because of” sex, and therefore violates the law, if the plaintiff can show that the decision was “tainted by awareness of sex . . . in any way,” even if legitimate reasons were at work as well.⁹² If sex was a factor in the decision, the defendant was liable for sex discrimination unless it could prove that it would have made the same decision even if sexism hadn’t been in play.

While Justice O’Connor agreed that Price Waterhouse discriminated against Hopkins because of her sex, she disagreed sharply with Justice Brennan’s reasoning. She, like the dissent, insisted that the plaintiff had to prove that sex was the “but for” cause of the decision. Justice O’Connor insisted that sexist “stray remarks,” “statements by nondecisionmakers” and even “statements by decisionmakers that were unrelated to the decisional process” were not enough to prove sex discrimination. This was merely sexism in the air.⁹³ But Justice O’Connor believed that Hopkins had established that the sexism at Price Waterhouse had come to earth and affected her candidacy for partnership:

It is as if Ann Hopkins were sitting in the hall outside the room where partnership decisions were being made. As the partners filed in to consider her candidacy, she heard several of them make sexist remarks in discussing her suitability for partnership. As the decisionmakers exited the room, she was *told* by one of those privy to the decisionmaking process that her gender was a major reason for the rejection of her partnership bid.⁹⁴

⁹¹ *Id.* at 241.

⁹² *Id.* at 230.

⁹³ *Id.* at 277 (O’Connor, J., concurring).

⁹⁴ *Id.* at 272.

The Justices thus articulated at least two distinct understandings of the “because of” requirement. On one view, a decision tainted by sexism constitutes discrimination “because of” sex only when the sexism itself would have been sufficient itself to account for the adverse outcome. This approach would require but-for causation. In the other view, discrimination occurs if sexism is a factor in the decision, even if the sexism was by itself insufficient to account for the decision. Subsequent to *Price Waterhouse*, the 1991 Amendments to Title VII codified the approach of Justice Brennan. The amendments clarified that a prohibited consideration need only be a “motivating factor” in a challenged decision in order for the decision to be declared discriminatory.⁹⁵

The question of unconscious bias does not play a role in any of these formulations. Neither the 1991 Act, the but-for requirement, nor Justice Brennan’s approach turns on the existence, or absence, of conscious bias. In *Price Waterhouse*, for example, it is irrelevant whether the sexist stereotypes that influenced the partners’ decision making were consciously held. Whether the partners would recognize that their views were objectionable, even whether they would recognize them as sex stereotypes, simply makes no difference. The significance of the fact that the partners openly expressed such stereotypes is not that the stereotypes were consciously held. Rather, the overt expression of such views merely performs an evidentiary function, removing any doubt as to whether such stereotypes were at work.

Hopkins would no doubt have found it much more difficult to prevail on her sex discrimination claim if the sex stereotypes were not openly expressed. In that case, one would have had to infer the existence of the relevant stereotypes. But, again, that inquiry would not in any way turn on whether the stereotypes were consciously or unconsciously held. If the partners had consciously decided to conceal their sex-stereotyped judgments of candidates for partnership (as many firms would undoubtedly do in the aftermath of the Court’s decision in *Price Waterhouse*), an aggrieved plaintiff would find it more difficult to prevail; that difficulty, however, would be no more or less daunting in a case of conscious concealment, compared to a case where the partners are genuinely unaware of having relied on such stereotypes.

Mixed motives might operate in another way as well. *Price Waterhouse* involved a large organization that made personnel decisions collectively.

⁹⁵ Civil Rights Act of 1991, Pub. L. No. 102-166, § 107(a), 105 Stat. 1071, 1075 (codified as amended at 42 U.S.C. § 2000e-(2)(5) (2000)) (adding § 703(m) to the Civil Rights Act of 1964).

Partnership review at Price Waterhouse didn't involve a simple decision by a personnel officer or supervisor; it was a multi-stage moon launch of a process in which every partner in the firm nationwide was invited to comment on the candidate.⁹⁶ The file was reviewed by two separate committees and finally submitted to the entire partnership for a vote. Hopkins was "held" by the second committee, which based its decision on the comments of 32 partners who commented on her candidacy—13 supported Hopkins, 8 opposed her, 8 hadn't formed an opinion, and 3 recommended she be "held."⁹⁷ Both supporters and detractors made sex-specific comments, such as that Hopkins should be more "lady-like," and both supporters and detractors made sex neutral criticisms, such as that Hopkins was abrasive and difficult to work with. In such a setting, it is impossible to identify a single decision maker, much less one with a discrete state of mind. Even if each individual acted for a single reason, motives might still be mixed in the sense that a variety of considerations—some legitimate, some not—accounted for the partnership's disposition of Hopkins's bid for partnership.

Despite Justice Brennan's extended disquisition on the subject, causation had almost nothing to do with the Price Waterhouse controversy. The real controversy involved the extent of the employer's duty to purify the workplace of prejudice. Justice Brennan and the liberals would presume discriminatory intent anytime bias had plausibly played a role in the challenged decision "in any way."⁹⁸ The effect of such a rule is to require employers to police the comments of their employees and punish or actively repudiate sexist comments. (Justice Brennan emphasized that Price Waterhouse "in no way disclaimed reliance on" the comments that reflected sex stereotypes.)⁹⁹

Justice O'Connor, by contrast, thought this went too far, turning Title VII into "thought control."¹⁰⁰ She would presume discriminatory intent only if sex stereotypes infected "the decisionmaking process"—which for Justice O'Connor was a discrete entity, separate from the work-a-day life of the firm, as suggested by her metaphor of a physically distinct corporate boardroom outside of which Hopkins waited.¹⁰¹ Justice O'Connor's rule would effectively require employers to keep sexism out of this metaphorical room. But despite

⁹⁶ *Price Waterhouse*, 490 U.S. at 232 (plurality opinion).

⁹⁷ *Id.* at 233.

⁹⁸ *See id.* at 239–42.

⁹⁹ *Id.* at 251.

¹⁰⁰ *See id.* at 262 (O'Connor, J., concurring in the judgment).

¹⁰¹ *See id.* at 273.

Justice O'Connor's demand for "but for" causation, this too has nothing to do with causation; instead, it—like Justice Brennan's opinion—defines the extent of an employer's duty to safeguard against discrimination.

The difference between Justices Brennan and O'Connor is simply that Justice O'Connor narrows the scope of the duty. For her, only comments made in connection with some discrete decision-making process count as proof of discriminatory intent. But many employees are the victims of bias, not in a formal decision-making process, but long before a formal review or promotion occurs. They suffer because supervisors give them "grunt work" instead of challenging assignments that offer the chance to impress and advance. They suffer subtle race- and sex-motivated slights and insinuations—what Justice O'Connor dismisses as "stray remarks"¹⁰²—that will harm their reputations. This type of discrimination, much more so than overt bigotry in a formal process, is a pervasive impediment to true equality of opportunity for women and racial minorities. Justice O'Connor downplays these types of bias, not because they can't cause injuries, but because employers can't eliminate them without intrusive policing of their employee's speech and perhaps even their thoughts. Her fear that civil rights law could become a form of state-imposed thought-control is perfectly valid. But it has nothing to do with distinguishing those cases in which sex bias "caused" a decision from those in which it did not. Instead, Justice O'Connor effectively defines discrimination pragmatically, in order to strike a balance between equal opportunity and freedom of expression.¹⁰³

3. *The Honest-Belief Rule*

Some commentators have argued that specific doctrinal rules developed by some of the federal circuits demonstrate a failure to appreciate the potential significance of unconscious bias.¹⁰⁴ The strongest argument for judicial

¹⁰² *Id.* at 277.

¹⁰³ This means that Title VII deliberately tolerates some discriminatory motivations. Justice O'Connor's pragmatic approach to antidiscrimination simply lets some discrimination go unaddressed because the costs of remedying it are too high. Viewed this way, O'Connor's approach may seem unsatisfying and even callous, a beady-eyed appraisal inappropriate to matters of civil rights. But perhaps the same pragmatism that informed her concurrence in *Price Waterhouse* also motivated her majority opinion upholding affirmative action in *Grutter v. Bollinger*, 539 U.S. 306 (2003). In a sense, the two positions are practical complements: given that the law cannot identify and prevent or punish every instance of invidious discrimination, affirmative action looks less like favoritism and more like a sensible social policy response to the discrimination that escapes the law's reach.

¹⁰⁴ *See, e.g.,* Krieger & Fiske, *supra* note 4, at 1029–38.

ignorance of unconscious bias cites the so-called “honest-belief rule.” In a nutshell, the rule says that if the employer can demonstrate that it “honestly believed” that it had good cause to take the challenged action, it defeats a claim of pretext in a single-motive case, even if the plaintiff can show that the employer in fact did not have good cause for the challenged adverse action. Professors Linda Krieger and Susan Fiske, two leading scholars of the role of psychological science in legal doctrine, observe that the Eleventh Circuit in *Jones v. Gerwens* held that “even if a Title VII claimant did not in fact commit the violation with which he is charged, an employer successfully rebuts any prima facie case of disparate treatment by showing that it honestly believed the employee committed the violation.”¹⁰⁵ In an otherwise sophisticated and insightful analysis, they argue that the honest-belief rule is

inconsistent with what empirical social psychologists have learned over the past twenty years [S]tereotypes can bias decision making *implicitly* by skewing the manner in which inherently ambiguous information about the stereotyped target is perceived [I]t is perfectly possible for a decision maker, whose biased judgment . . . caused him to discriminate . . . to believe that his judgment and resulting decision were based entirely on legitimate nondiscriminatory reasons.¹⁰⁶

Thus, Professors Krieger and Fiske oppose the honest-belief rule.

We share their opposition to the honest-belief rule, but the rule is undesirable for reasons that have nothing to do with a factual question of unconscious motivation. Along with Supreme Court Title VII jurisprudence more generally, the honest-belief rule does not, in fact, preclude consideration of unconscious bias. Instead, the honest-belief rule is a contestable, but reasonable extension of the formal single-motive proof structure. The rule does not make liability contingent on whether the defendant consciously believed it discriminated. It makes liability contingent on whether the plaintiff can show that the defendant’s proffered rationale was a pretext—a sham or cover for discriminatory intent. If the defendant believed the rationale, and if the rationale would justify the challenged action, then the rationale, if mistaken, is not a pretext—it’s simply a mistake. An easy case: the defendant fires the plaintiff based on the mistaken belief that the plaintiff has been embezzling company funds. At trial it is revealed that the plaintiff has not embezzled funds but instead that there was an accounting error. Although the

¹⁰⁵ *Id.* at 1036 (citing *Jones v. Gerwens*, 874 F.2d 1534, 1540 (11th Cir. 1989)).

¹⁰⁶ *Id.* (footnote omitted).

defendant thus did not have cause to fire the plaintiff, the honest-belief rule would justify a directed verdict for the defendant: the fact-finding demonstrated that the defendant was motivated by a legitimate concern, namely embezzlement. The challenged decision was not in fact made because of race—it was made because of a mistaken belief that the employee had embezzled funds. People can be motivated sincerely by concerns that turn out to be unfounded; as long as the defendant believed the concern was valid when it made the decision, the plaintiff can't prevail in a single-motive case because the plaintiff has not established discrimination by process of elimination—it has not eliminated the defendant's alleged motivation for the challenged action.

The real importance of the honest-belief rule is to preserve the default of at-will employment; the plaintiff will not prevail when termination results from a mistake rather than from good cause. The status of at-will employment is, as we have suggested, a central ideological conflict in employment discrimination law, and it, more than any factual observation about bias, drives the legal analysis. In the absence of the honest-belief rule, Title VII would amount to a sort of weak good cause regime for plaintiffs who prove the prima facie *McDonnell Douglas* case; if the plaintiff can prove that she should not have been fired—either because the defendant was mistaken or biased—then the opponent of the honest-belief rule wants the plaintiff to prevail. By contrast, the honest-belief rule preserves at-will employment: if the defendant honestly but mistakenly believed the nondiscriminatory reason, it has not discriminated on the basis of race and the plaintiff is simply a victim of a termination without good cause, which is unfortunate but not actionable.

Of course, many cases will involve factual ambiguities: was the defendant's honest belief influenced in some way by the prohibited motivations? This is no doubt the concern that leads Krieger and Fiske to insist that the rule is inconsistent with contemporary psychological science. But their concern goes to the question of what we mean by an "honest belief." Nothing in the rule stipulates that an honest belief can refer only to the decision maker's conscious motivations. Indeed, how could a fact finder know what a decision maker's conscious motivations are? In effect, the honest-belief rule requires the fact finder simply to assess a given rationale in light of the evidence available to the decision maker at the time the decision was made. It asks: would a reasonable decision maker have considered this sufficient justification for the challenged action given what the decision maker knew at the time? The rule seeks to exclude the hindsight bias that would be

introduced if the fact finder were to evaluate the proffered rationale in light of the subsequent revelation that the rationale is mistaken. But this does not involve any limitation of “honest belief” to conscious psychological processes, nor could it because fact finders do not have direct evidence of mental state.

The honest-belief rule also attempts to enforce the evidentiary requirements for a mixed-motive case. Mixed-motive cases involve two potential motivations: first, the mistaken belief that, if true, would justify the challenged action and, second, race or sex. A motive based on a factual mistake is no different than a motive based on accurate information: the question is whether that motive in fact caused the challenged action or whether a prohibited motive did. When the honest-belief rule was developed, most circuits followed Justice O’Connor’s concurrence in *Price Waterhouse*¹⁰⁷ and required direct evidence of discriminatory motivation in order to convert a single-motive case into a mixed-motive case.¹⁰⁸ The honest-belief rule was a means of precluding a mixed-motive jury instruction in the absence of direct evidence. The plaintiff’s evidence was subject to the honest-belief rule precisely because it had not adduced direct evidence and, therefore, the *McDonnell Douglas* deductive structure was thought to be the only evidentiary path available.¹⁰⁹ (The Supreme Court repudiated the direct-evidence requirement for mixed-motive analysis in *Desert Palace, Inc. v. Costa*¹¹⁰ in 2003, and it is not clear that the honest-belief rule should have much continuing relevance post *Costa*.) In sum, the honest-belief rule does not implicate the distinction between conscious and unconscious motivations, but it did prevent plaintiffs from alleging a mixed-motive theory without direct evidence of discrimination.

B. Constitutional Law: Equal Protection

Just as Title VII draws no distinction between covert and unconscious bias, so too does the covert–unconscious distinction play no role in courts’ adjudications of disputes under the Equal Protection Clause. Nonetheless, it is

¹⁰⁷ *Price Waterhouse v. Hopkins*, 490 U.S. 228, 275 (1989) (O’Connor, J., concurring) (arguing that in mixed-motive employment discrimination cases where the plaintiff has presented direct evidence to show that an “illegitimate criterion” was a “substantial factor” in an employment action, the burden of persuasion should be shifted to the defendant).

¹⁰⁸ See, e.g., Michael J. Zimmer, *The New Discrimination Law: Price Waterhouse Is Dead, Whither McDonnell Douglas?*, 53 EMORY L.J. 1887, 1910 (2004).

¹⁰⁹ See *supra* Part II.A.1 (explaining the *McDonnell Douglas* single-motive burden of proof framework).

¹¹⁰ 539 U.S. 90, 101 (2003). The Court interpreted a 1991 congressional amendment to Title VII to require a plaintiff to present “sufficient evidence for a reasonable jury to conclude, by a preponderance of the evidence” that a protected class was a “motivating factor” in an employment practice. *Id.*

easy to understand the belief that constitutional doctrine is somehow in tension with the possibility of unconscious bias. The Supreme Court's constitutional race discrimination jurisprudence is replete with references to "purpose" and "intent."¹¹¹ In its 1976 decision in *Washington v. Davis*, the Court adopted the so-called "discriminatory purpose" test.¹¹² In deciding an equal protection challenge to a personnel test employed by the District of Columbia police department that failed four times as many black applicants as white applicants, the Court ruled that such a claim of racial discrimination requires a finding of discriminatory intent.¹¹³ The Court stated that "a purpose to discriminate must be present."¹¹⁴ During the next few years, the Court twice reiterated the so-called intent requirement, stating forthrightly that "proof of racially discriminatory intent or purpose is required to show a violation of the Equal Protection Clause."¹¹⁵ Lower federal courts followed the Supreme Court's lead, often speaking in terms of discriminatory purpose.¹¹⁶

It would be a mistake, though, to read *Washington v. Davis* and its progeny as exempting from prohibition discrimination prompted by unconscious bias. The Court's use of the term discriminatory purpose was fortuitous and not plausibly intended to apply only to intentional discrimination. The Court has used the terms "motive" and "purpose" interchangeably, as evidenced, for example, by the Court's observation that a challenged decision is unconstitutional if discriminatory purpose is a "motivating factor."¹¹⁷ In adopting the discriminatory purpose test, the Court was not discounting the significance of unconscious bias, so much as it was choosing among a range of plausible formulations of the nondiscrimination mandate in the aftermath of the invalidation of Jim Crow.

¹¹¹ See, e.g., *Washington v. Davis*, 426 U.S. 229 (1976).

¹¹² See *id.* at 241 (holding that a policy that has a racially disparate impact without a racially discriminatory purpose does not violate the Equal Protection Clause).

¹¹³ See *id.* at 242 (explaining that while disparate impact is a factor in showing invidious racial intent, it is not dispositive).

¹¹⁴ *Id.* at 239 (quoting *Akins v. Texas*, 325 U.S. 398, 403–04 (1945)).

¹¹⁵ *Vill. of Arlington Heights v. Metro. Hous. Dev. Corp.*, 429 U.S. 252, 265 (1977).

¹¹⁶ See, e.g., *Brown v. City of Oneonta*, 221 F.3d 329, 337 (2d Cir. 1999) (stating that a "plaintiff could also allege that a facially neutral statute or policy has an adverse effect and that it was motivated by discriminatory animus," to establish equal protection violation); *United States v. Galloway*, 951 F.2d 64, 65 (5th Cir. 1992) (citing *Personnel Adm'r v. Feeney*, 442 U.S. 256, 279 (1979)) ("Discriminatory purpose in an equal protection context implies that the decisionmaker selected a particular course of action at least in part because of, and not simply in spite of, the adverse impact it would have on an identifiable group.").

¹¹⁷ *Arlington Heights*, 429 U.S. at 265.

Washington v. Davis presented the question of the constitutional standard applicable to formally race-neutral laws that are challenged as racially discriminatory.¹¹⁸ The appeals court had invalidated the challenged personnel test on the basis of its disparate impact on African-American applicants to the police force.¹¹⁹ The disparate impact standard was applicable to employers under Title VII,¹²⁰ and the lower court had embraced it as the constitutional standard as well.¹²¹ According to this approach, any practice that disproportionately burdened a historically disadvantaged racial group would need to satisfy a burden of justification in order to escape invalidation.

The Supreme Court reversed, rejecting disparate impact as the constitutional standard.¹²² Disparate impact may constitute illegal discrimination under a federal statute such as Title VII, the Court observed, but a constitutional claim requires something more. That “something more” the Court referred to as discriminatory purpose. The Court emphasized that “our cases have not embraced the proposition that a law or other official act, without regard to whether it reflects a racially discriminatory purpose is unconstitutional solely because it has a racially disproportionate impact.”¹²³ The Court objected to an impact test because such an approach “would be far-reaching and would raise serious questions about, and perhaps invalidate, a whole range of tax, welfare, public service, regulatory, and licensing statutes that may be more burdensome to the poor and to the average black than to the more affluent white.”¹²⁴ In adopting the discriminatory purpose test, the Court sought to avoid putting into question the constitutionality of such statutes.¹²⁵ Indeed, a disparate impact test could have placed a heavy burden of justification on practically any governmental practice. Faced with that possibility, even Justices Marshall and Brennan, the most liberal members of the Court, declined to push for disparate impact as the constitutional standard.

¹¹⁸ See *Davis*, 426 U.S. at 233 (evaluating whether a personnel test that disproportionately excluded black applicants from jobs with a police department violated the Equal Protection Clause).

¹¹⁹ See *id.* at 237 (explaining the court of appeals’s holding that discriminatory impact without discriminatory purpose is sufficient for a constitutional claim).

¹²⁰ The controversy arose before the passage of Title VII, and the complaint had not been amended to state a claim under Title VII.

¹²¹ See *id.* at 238 (explicitly rejecting the court of appeals’s application of the Title VII standard).

¹²² See *id.* at 239 (“We have never held that the constitutional standard for adjudicating claims of invidious racial discrimination is identical to Title VII, and we decline to do so today.”).

¹²³ *Id.* at 239.

¹²⁴ *Id.* at 248.

¹²⁵ *Id.*

In articulating the discriminatory purpose standard, the Court also rejected another plausible formulation of the nondiscrimination mandate that was in play at the time it decided *Washington v. Davis*. Five years prior to *Washington v. Davis*, the Court had decided *Palmer v. Thompson*.¹²⁶ In that case, the Court upheld the decision of city officials in Jackson, Mississippi, to close the city's swimming pools rather than integrate them in response to a court order that prohibited official segregation of public facilities.¹²⁷ City officials integrated other public facilities but decided to close the pools in order to, according to the city, maintain peace and save money, having concluded that the pools could not be operated safely and economically on an integrated basis.¹²⁸ In upholding the closure, the Court reasoned that these legitimate justifications could not "be impeached by demonstrating that racially invidious motivations had prompted the city council's action."¹²⁹ The Court's decision in *Palmer* could have provided the basis for a rigidly formal approach to nondiscrimination, in which a challenged practice would be invalidated if, and only if, it formally discriminated on the basis of race. An equal protection jurisprudence based on the Court's decision in *Palmer* could have precluded any inquiry into legislative motivation.

In *Washington v. Davis*, the Court rejected the principle of *Palmer*.¹³⁰ The Court's adoption of the discriminatory purpose standard made abundantly clear that inquiry into motivation is permissible in evaluating formally neutral practices that are challenged as racially discriminatory.¹³¹ The Court rejected an approach that would have precluded inquiry into motivation, and thus permitted challenges to laws that are not formally discriminatory.

This background makes clear that the discriminatory purpose standard was not a rejection of unconscious bias claims. The Court seized on discriminatory purpose as a means of rejecting two alternative formulations of the nondiscrimination mandate. The disparate impact approach potentially could have put into question the constitutionality of a wide range of governmental practices, while the formal approach of the *Palmer* Court would have insulated from review even those formally neutral practices that seemed intended to

¹²⁶ 403 U.S. 217 (1971).

¹²⁷ *Id.*

¹²⁸ *See id.* at 219 (providing background information and the procedural posture of the case).

¹²⁹ *Davis*, 426 U.S. at 242–43.

¹³⁰ *Id.* at 234 (indicating that *Palmer* is not applicable to *Washington v. Davis*).

¹³¹ *Id.* at 242 ("[I]nvidious discriminatory purpose may often be inferred from the totality of the relevant facts.").

function as did the Jim Crow laws of the pre-*Brown* era. The discriminatory purpose standard might be viewed as a middle ground between these two extremes. In rejecting an impact standard, the Court narrowed the range of impermissible practices, but in allowing inquiry into motivation, the Court made clear that formally race-neutral practices could be subject to challenge as well. In sum, then, the Court used the language of discriminatory intent and discriminatory purpose not to distinguish conscious from unconscious bias, but instead to distinguish the constitutional standard from both the disparate impact standard of Title VII and the formal approach of the *Palmer* Court.

Just as the choice of the discriminatory purpose standard did not reflect any desire to exclude decisions prompted by unconscious bias from the nondiscrimination mandate, so too do the evidentiary considerations identified by the Court not distinguish between conscious and unconscious influences. The year after its decision in *Washington v. Davis*, the Court decided *Village of Arlington Heights v. Metropolitan Housing Authority*,¹³² in which it set forth a framework for proving discriminatory purpose. The proof process was less formal than that applicable in the Title VII context, but as with Title VII, the Court's approach did not invoke any distinction between conscious and unconscious bias. *Arlington Heights* concerned a challenge to a suburban municipality's denial of a rezoning application for a housing developer who planned to develop a mixed income, and racially integrated, housing complex in a predominantly white town north of Chicago.¹³³ To construct the townhouse complex, the developer needed the land rezoned from single-family residential to multi-family.¹³⁴ After a number of controversial public hearings, the town denied the developer's rezoning request.¹³⁵ The town's official justification was that the development "threatened to cause a measurable drop in property value."¹³⁶

The Court used the case as an opportunity to set forth a number of factors for lower courts to consider in deciding whether a formally neutral practice challenged as racially discriminatory violates the Equal Protection Clause. The Court noted that the inquiry would begin by ascertaining whether the challenged decision "bears more heavily on one race than another."¹³⁷ The

¹³² 429 U.S. 252 (1977).

¹³³ *Id.* at 255.

¹³⁴ *Id.* at 257.

¹³⁵ *Id.* at 257-59.

¹³⁶ *Id.* at 258.

¹³⁷ *Id.* at 266 (quoting *Washington v. Davis*, 426 U.S. 229, 242 (1976)).

Court reasoned that if “a clear pattern, unexplainable on grounds other than race, emerges from the effect of the state action . . . [t]he evidentiary inquiry is then relatively easy.”¹³⁸ If no nonracial account of the effect of the challenged practice seems plausible, then the court would conclude that the practice is racially discriminatory. Although such cases are no doubt quite rare, what is noteworthy is that the Court’s discussion of the evidentiary inquiry makes no mention of the state of mind of any decision maker. Rather, the Court centers the inquiry on an examination of what actually happened.

In the vast majority of cases, of course, the challenged action will be subject to alternative explanations, in which case the court should consider the historical background of the decision as well as “the specific sequence of events leading up to the challenged decision.”¹³⁹ A departure from normal procedures or considerations, for example, might justify a finding that the challenged action was discriminatory, especially if the usual procedures or considerations would have weighed heavily in favor of an alternative decision.¹⁴⁰ The Court noted that the evidentiary examination of the history of the decision should include meeting minutes and reports produced in connection with the decision.¹⁴¹ Again, these evidentiary considerations do not distinguish between conscious and unconscious bias.¹⁴² One might conclude that the plaintiff need not prove bias at all, but instead simply that the decision would have been different but for the races of the parties. As the Court noted in another post-*Washington v. Davis* decision, *Personnel Administrator of Massachusetts v. Feeney*,¹⁴³ discriminatory purpose refers to whether an action was taken “at least in part ‘because of,’ not merely ‘in spite of,’ its adverse effects upon an identifiable group.”¹⁴⁴

¹³⁸ *Id.*

¹³⁹ *Id.* at 267.

¹⁴⁰ *See id.* (explaining that, if the land at issue in Arlington Heights was rezoned in response to the developer’s plan, that would be considered a significant departure from normal procedure).

¹⁴¹ *See id.* at 268 (providing examples of the type of “legislative or administrative history” that may be relevant).

¹⁴² The more interesting issue raised in *Arlington Heights* concerns the meaning of discriminatory purpose or motivation. The Court noted that during one of the well-attended public meetings of the Planning Commission, a number of community members mentioned the “social issue” of the “desirability . . . of low- and moderate-income housing that would probably be racially integrated.” *Id.* at 257–58. The district court concluded, after a trial, that the town denied the permit due to concerns about a decline in property values. But what if property values were expected to decline at least partly because of the racially integrated nature of the project? What if property values would in fact decline more as a result of a racially integrated complex than an all-white complex, and the city’s calculus reflected that unfortunate fact?

¹⁴³ 442 U.S. 256 (1979).

¹⁴⁴ *Id.* at 279.

One might conclude that the underlying concern that has prompted so much scholarly criticism of the discriminatory purpose test is not so much its silence with respect to unconscious bias, but rather the near impossibility of invalidating legislative acts that are formally race neutral. One criticism, then, is that so long as a practice is formally race neutral, it will be extraordinarily difficult to satisfy the burden of proof. But that difficulty implicates the distinction between overt and covert discrimination, not the distinction between conscious and unconscious bias. Consider the meeting minutes and committee reports that a court would consider in deciding whether a subsequent decision was discriminatory. If the meeting minutes recounted racially bigoted remarks made in connection with the challenged decision, it would be easy for a court to conclude that the resulting decision was discriminatory. But suppose, as is almost always the case, that the decision makers are savvy enough not to have their biases entered into the official record. Then, even if every member of the city council in *Arlington Heights*, for example, voted to deny the rezoning application solely because they consciously opposed racially integrated housing, it would still be difficult to prove discriminatory purpose. Such a covert, yet conscious, discriminatory purpose would be no easier to prove than a decision motivated by unconscious bias. In both cases, the problem would be the lack of evidence. Cases of unconscious bias will involve little direct evidence of discrimination, but so too will almost all cases involving conscious yet covert bias. Discrimination will be difficult to prove whenever bias is covert.

A potent criticism of the discriminatory purpose standard, then, is that while in theory it holds out the possibility of invalidating formally race-neutral decisions when they are animated by a discriminatory purpose, in reality, the hurdle of proving discriminatory purpose is so daunting that virtually no claim will surmount it. Challenges to many arguably unjust laws have floundered on the challenge of demonstrating discriminatory intent. The crack cocaine versus powdered cocaine sentence disparity, felon disenfranchisement laws, the administration of the death penalty—these are areas where constitutional challenges that the policies are racially discriminatory have fallen flat, unable to satisfy the burden of proof.¹⁴⁵

The emphasis on unconscious bias is a means of forcing a relaxation of the burden of proof. Some commentators may hope that a widespread embrace of

¹⁴⁵ See, e.g., *McCleskey v. Kemp*, 481 U.S. 279 (1987); *Cotton v. Fordice*, 157 F.3d 388 (5th Cir. 1998); *United States v. Clary*, 34 F.3d 709 (8th Cir. 1994).

unconscious bias theory would cause many cases to come out differently. That hope reflects the intuition that a more capacious understanding of the causes of discrimination might prompt, or at least enable, fact finders to designate more decisions and policies as racially discriminatory. If one believes that discrimination motivated by unconscious bias is pervasive, then one might become more willing to infer discrimination on the basis of ambiguous evidence, even without the sort of statements typically associated with conscious bias. Again, though, the important distinction here is not between conscious and unconscious bias—believing that covert discrimination is pervasive could also incline one to infer discrimination from ambiguous evidence.

Judicial disinclination to designate challenged practices as racially discriminatory also reflects legitimate, if not always persuasive, concerns about remedy. Consider the most well-known case concerning the death penalty, *McCleskey v Kemp*.¹⁴⁶ In that case, the Court confronted rigorous statistical studies that demonstrated the importance of race in the administration of the death penalty.¹⁴⁷ The so-called Baldus study showed that the demographic combination of a black defendant, such as McCleskey, and a white murder victim was the most likely to result in a death sentence.¹⁴⁸ However, the Court rejected McCleskey's claim, and declined to overturn his death penalty.¹⁴⁹

The problem, as the Court saw it, centered on the lack of evidence that racial bias had entered into official decision making in the specific case of Warren McCleskey. According to this reasoning, even if statistics incontrovertibly showed that in the State of Georgia the likelihood of a defendant's receiving the death penalty depended on his race and that of his victim, a claim of racial discrimination would fail absent evidence that race had tainted the sentence received by that particular defendant.

¹⁴⁶ 481 U.S. 279.

¹⁴⁷ See *id.* at 283–84 (presenting the issue of the case as whether a statistical analysis showing that race affects sentencing can render the defendant's capital sentence unconstitutional).

¹⁴⁸ See *id.* at 286–88 (examining David C. Baldus et al., *Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience*, 74 J. CRIM. L. & CRIMINOLOGY 661 (1983)). Also, this conclusion applied to what Baldus described as cases where sometimes the death penalty was imposed and sometimes not. In the most serious category of cases, the death penalty was regularly imposed, irrespective of the race of the parties.

¹⁴⁹ See *id.* at 319 (holding that the law was correctly applied with respect to McCleskey).

Commentators have justifiably criticized the Court's decision in *McCleskey* as erecting a nearly insurmountable burden of proof.¹⁵⁰ Under the Court's approach, *McCleskey* could only prevail in his discrimination claim with smoking-gun evidence, such as racially bigoted remarks by a prosecutor or juror, which rarely emerges. The case thus might seem a perfect opportunity to criticize the discriminatory purpose standard that governed *McCleskey*'s equal protection challenge.

In fact, though, the problem in *McCleskey* is not one of proof. *McCleskey* could well have met the burden of proof—preponderance of the evidence—without any evidence that the specific prosecutor or jury in his case had made its decisions on the basis of race. Purely statistical evidence could have sufficed. Imagine, controlling for all nonracial factors, that the death penalty was given to 95% of black defendants with a particular profile, but only 10% of white defendants with that same profile. With such statistics, someone in *McCleskey*'s position could credibly argue that he more likely than not would have been spared the death sentence had he been white.

The real issue in *McCleskey*, and the reason the Court focuses on evidence relating to his particular case, is remedial. Evidence of discrimination that was specific to his case—for example, a prosecutor's racially bigoted remarks in discussing his decision to seek the death penalty—would have been useful because it would have allowed the Court to limit the remedial order to that particular case. The Court, for example, could have invalidated the death penalty in *McCleskey*'s case without imperiling the death sentences accorded in other cases. In contrast, a decision to invalidate *McCleskey*'s death sentence based on statistical evidence would have had far-reaching implications. If the Court had invalidated *McCleskey*'s sentence because it involved the racial combination—white victim, black defendant—most likely to result in a death sentence, then all other similarly situated defendants would certainly be entitled to have their death sentences overturned as well. And once those death sentences were invalidated, courts would hear challenges from those defendants in the next highest risk category for receiving a death sentence. And so on and so on. A remedial order in favor of *McCleskey* based on statistical evidence would thus push toward abolition of the death penalty,

¹⁵⁰ See, e.g., Jeffrey Fagan & Mukul Bakhshi, *New Frameworks for Racial Equality in the Criminal Law*, 39 COLUM. HUM. RTS. L. REV. 1, 3 (2007) (“[A]bsent some type of indiscreet comment by a prosecutor or judge that provides ‘smoking-gun’ proof of discriminatory intent, the evidentiary threshold required by *McCleskey* was then and still is virtually insurmountable.”); Randall L. Kennedy, *McCleskey v. Kemp: Race, Capital Punishment, and the Supreme Court*, 101 HARV. L. REV. 1388, 1404–05 (1988).

unless the Court was willing to wage war on the discretion that produces racial disparities in death sentencing. The *McCleskey* Court did not want, in effect, to prohibit the death penalty, nor to attempt to eliminate discretion in death sentencing.¹⁵¹ Of course, this simplified scenario does not exhaust the remedial possibilities available to the *McCleskey* Court, but it does accurately suggest the difficulties the Court would have confronted in basing a decision on statistical evidence of discrimination. Avoiding either the elimination of the death penalty or the elimination of discretion would have likely engaged the Court in more extensive and ongoing management of the sentencing process than it would have preferred.

This account of *McCleskey* highlights the irrelevance of awareness of unconscious bias in the decisional calculus. The Court's decision in *McCleskey* has been extensively criticized, justifiably so in our view.¹⁵² But it misconstrues the issues in the case to imagine that greater judicial awareness of unconscious bias would, or should, have changed the outcome. The problem with the Court's decision was not that it relied on a narrow understanding of discrimination as conscious and overt, but rather that it failed to craft a creative solution to what was an unquestionably difficult problem. At the very least, the Court should not have permitted its understandable concerns about remedy to circumscribe its recognition of discrimination.

Analogous anxieties about the judicial role underlie the Court's disinclination to strike down formally race-neutral laws in other areas. Consider, for example, the consistently criticized disparity in sentencing for crack versus powdered cocaine, whereby possession of crack cocaine is punished much more harshly than possession of an equivalent amount of powdered cocaine.¹⁵³ The formally race-neutral sentencing scheme results in African-American drug defendants receiving much harsher penalties than white defendants, because African-American drug defendants are more likely than white defendants to be prosecuted for a crack cocaine offense, whereas white defendants are more likely to have powdered cocaine. Courts' disinclination to strike down this sentencing disparity as racially discriminatory has little to do with judicial acknowledgement of unconscious bias and everything to do with a sensible respect for the outcomes of the political

¹⁵¹ And, of course, even if the Court did eliminate sentencing discretion, prosecutorial discretion would remain and would be practically impossible to eliminate.

¹⁵² See, e.g., Symposium, *Pursuing Racial Fairness in Criminal Justice: Twenty Years After McCleskey v. Kemp*, 39 COLUM. HUM. RTS. L. REV. 1 (2007).

¹⁵³ See, e.g., David A. Sklansky, *Cocaine, Race, and Equal Protection*, 47 STAN. L. REV. 1283 (1995).

process. Courts will almost never strike down, as racially discriminatory, validly enacted legislation that is formally race neutral. Judicial acceptance of the idea of unconscious bias would not make the invalidation of such laws any more likely.

To take an even more extreme, and timely, example, consider immigration law. Suppose that Congress passed an extraordinarily harsh immigration law in response to many voters' conscious dislike of immigrants from Mexico. If the law was formally race neutral, a court would be unlikely to strike it down, even if everyone agreed that racially discriminatory motivations underlie it. If even evidence of such conscious bias would not result in the invalidation of the law, then it is difficult to imagine that greater awareness of unconscious bias would. The legal decision maker's awareness of unconscious bias is largely irrelevant to these controversies. In the sorts of cases where a court is unlikely to strike down a challenged practice, greater awareness of unconscious bias would not cause it to do so.

In cases where remedial concerns or anxieties about judicial role do not loom so large, courts are more willing to strike down decisions that are challenged as racially discriminatory and to see little reason to sort out whether the challenged decision stemmed from unconscious rather than conscious discriminatory motivations. Consider peremptory challenges.¹⁵⁴ Courts may strike down challenges as impermissibly race-based, irrespective of a lawyer's claim that he or she did not intend to rely on race. Almost by definition, the lawyer whose strike is challenged will always say that some reason other than race explains the strike. In deciding whether to permit the strike, courts rarely, if ever, founder on the question of intent.¹⁵⁵ There is no exemption for strikes that are discriminatory, but not intentionally so. The distinction between conscious and unconscious reliance on race plays a small role in the determination of whether to allow the peremptory strike of a particular juror.¹⁵⁶

¹⁵⁴ See generally *Batson v. Kentucky*, 476 U.S. 79 (1986) (holding that prosecutors cannot challenge the selection of a juror based solely on the juror's race).

¹⁵⁵ See, e.g., Sheila Foster, *Intent and Incoherence*, 72 TUL. L. REV. 1065, 1094–97 (1998) (explaining that race-based peremptory challenges may be struck down even without proof of conscious intent to discriminate, and concluding that “[m]ore often than not, a finding of discriminatory purpose is made regardless of the actor’s state of consciousness”).

¹⁵⁶ Yet, the rhetoric of unconscious bias may make a judge more willing to strike down a practice, such as a peremptory challenge, by allowing the judge to do so without calling the decision maker a liar. A judge that rejects a lawyer's stated rationale as false would be implying that the lawyer lied, unless the court attributes the discriminatory practice to unconscious bias.

In sum, then, greater awareness of unconscious bias would not prompt courts to strike down practices that, for a variety of reasons, they don't want to strike down. Nor does the lack of doctrinal attention accorded unconscious bias block courts from invalidating discriminatory decisions that they want to strike down, even those that may have resulted from unconscious bias.

C. Disparate Impact, Affirmative Action, and the Burden of Proof

Claims of unconscious bias are also used to support affirmative action and disparate impact—two embattled features of our antidiscrimination regime¹⁵⁷—by situating them as extensions of the nondiscrimination mandate rather than alternatives to it. But, in fact, their characterization as consistent with the goal of nondiscrimination does not turn on the existence of unconscious bias. Rather, the pivotal issue concerns the extent of racial discrimination, notwithstanding its prohibition. If racial discrimination remains widespread, then both affirmative action and disparate impact seem sensible responses to our failure to eliminate discrimination, indirect means of realizing a central goal of antidiscrimination law. Alternatively, if racial discrimination is insignificant, then each policy seems to violate the nondiscrimination mandate in the name of social engineering, perhaps as a result of racial group politics.

The potential clash between affirmative action and nondiscrimination is sharpest.¹⁵⁸ Affirmative action may entail precisely what the nondiscrimination mandate prohibits: treating individuals differently on

¹⁵⁷ The Supreme Court has placed severe restrictions on affirmative action by colleges and universities and has practically prohibited race-based student assignments in primary and secondary schools. See *Parents Involved in Cmty. Sch. v. Seattle Sch. Dist. No. 1*, 551 U.S. 701 (2007); *Grutter v. Bollinger*, 539 U.S. 306 (2003); *Gratz v. Bollinger*, 539 U.S. 244 (2003). The Court has also limited the availability of disparate impact claims under federal statutes. In revising what had seemed to many a settled understanding of the doctrine, the Court ruled in *Alexander v. Sandoval*, 532 U.S. 275 (2001), that there is no private right of action to enforce the regulations promoting disparate impact under Title VI, the federal statute that prohibits discrimination by all entities receiving federal funds. The disparate impact prohibition could, in the aftermath of *Sandoval*, be enforced by regulatory agencies, but not by aggrieved individuals through litigation. A disparate impact rule continues to be a part of the federal law prohibiting employment discrimination, but its application continues to be a topic of great disagreement. Affirmative action and disparate impact are also limited in the Title VII context.

¹⁵⁸ See, e.g., *Grutter*, 539 U.S. at 326–27 (discussing the Court's standard of review for race-based affirmative action policies); *Adarand Constructors, Inc. v. Peña*, 515 U.S. 200, 235 (1995) (holding that the standard of strict scrutiny governs whether race-based classifications violate the equal protection component of the Fifth Amendment's Due Process Clause); *City of Richmond v. J.A. Croson*, 488 U.S. 469, 498–99 (1989) (holding that a "generalized assertion" of past racial discrimination cannot justify "rigid" racial quotas for the awarding of public contracts).

account of race.¹⁵⁹ Far from prohibiting discrimination, affirmative action is, strictly speaking, a form of discrimination—an effort to manipulate group outcomes by treating individuals differently on the basis of race.

Disparate impact does not directly contravene the nondiscrimination mandate, but it does extend beyond that core prohibition in that decision makers may be found liable even when they genuinely have not taken race into account.¹⁶⁰ Rather than a narrowly targeted means of eliminating discrimination, a disparate impact rule could be criticized as a means of social engineering.¹⁶¹ This criticism is understandable, to the extent that the application of a disparate impact rule would depend on some underlying judgment about which patterns of group outcomes are acceptable and which are not. Moreover, in threatening the invalidation of practices on the basis of their racial consequences, disparate impact encourages decision makers to consciously consider the racial consequences of their decisions. If a decision maker covered by a disparate impact rule does not want to violate the law, she or he would do well to evaluate how any particular policy affects different groups. The prospect of liability under disparate impact theory thus encourages decision makers to do exactly what the nondiscrimination mandate prohibits them from doing.

The unconscious bias discourse blunts these criticisms of affirmative action and disparate impact, by making each seem continuous with the opposition to discrimination. Unconscious bias theory implies that racial discrimination remains widespread, albeit in a more subtle form than in generations past. The persistence of subtle discrimination, in turn, weighs in favor of the relaxation of the burden of proof as courts attempt to ferret out discrimination. If unconscious bias predictably leads to discriminatory practices even when decision makers do not seem to have considered race, then requiring some sort of direct evidence that race played a role would immunize many discriminatory policies. Relax the burden of proof enough and the prohibition on treating individuals differently on account of race transmutes into a disparate impact rule. Racially skewed outcomes alone become potentially persuasive evidence

¹⁵⁹ *Grutter*, 539 U.S. at 326 (quoting *Adarand*, 515 U.S. at 227) (“[G]overnment may treat people differently because of their race only for the most compelling reasons.”).

¹⁶⁰ *See, e.g.*, *Griggs v. Duke Power Co.*, 401 U.S. 424, 431 (1971).

¹⁶¹ *See, e.g.*, *Washington v. Davis*, 426 U.S. 229, 248 (1976) (“A rule that a statute designed to serve neutral ends is nevertheless invalid, absent compelling justification, if in practice it benefits or burdens one race more than another would be far-reaching and would raise serious questions about, and perhaps invalidate, a whole range of tax, welfare, public service, regulatory, and licensing statutes that may be more burdensome to the poor and to the average black than to the more affluent white.”).

of discrimination. On this account, then, disparate impact doctrine is needed to police discriminatory practices that, under a higher burden of proof, could not be proven to be discriminatory. Similarly, the prospect of unconscious bias depicts affirmative action as a means of countering current discrimination, an effort to even the playing field rather than provide an advantage on the basis of race. If discrimination is pervasive, yet nearly impossible to eliminate directly because it is both subtle and unintentional, then some form of affirmative action becomes a sensible means of leveling what would otherwise be a very unfair playing field for those groups who are the object of implicit bias.

In sum, then, unconscious bias has been used to depict affirmative action and disparate impact alike as desirable, indeed necessary, means of countering racial discrimination that is continuing and pervasive.¹⁶² Rather than contradict or go beyond an individualistic nondiscrimination mandate, they become indirect means of securing it. Affirmative action counters the discrimination that we know occurs but cannot identify, while disparate impact prohibits those practices whose racially skewed burdens suggest that unconscious bias may have played a role.¹⁶³ This characterization of the justification for affirmative action and disparate impact makes each policy less ideologically contestable than they might otherwise appear to be. If one believes that unconscious bias is pervasive, then one's opposition to such discrimination could itself be sufficient to prompt support of affirmative action and a disparate impact rule.¹⁶⁴

However, affirmative action and disparate impact might be situated as extensions of the nondiscrimination mandate without any reliance on the idea of unconscious bias. Rather, the core claim is that widespread discrimination persists and is so subtle that it cannot be precisely identified and eliminated.

¹⁶² See, e.g., Michael Selmi, *Testing for Equality: Merit, Efficiency, and the Affirmative Action Debate*, 42 UCLA L. REV. 1251 (1995) (arguing that affirmative action can prevent unconscious discrimination in the workplace); Michael J. Yelnosky, *The Prevention Justification for Affirmative Action*, 64 OHIO ST. L.J. 1385 (2003) (same).

¹⁶³ More generally, the unconscious bias research thus encourages us all to incorporate into our thinking a sort of disparate impact norm, to be suspicious of outcomes that disadvantage minorities. Even more, the belief that one is infected with unconscious bias may prompt one to engage in affirmative action (in the sense of according some positive weight to minorities' race in a selection process), understood, again, not as a form of social engineering, but as a prophylactic, a means of preventing the discrimination that unconscious bias would otherwise cause.

¹⁶⁴ See Greenwald & Krieger, *supra* note 4, at 961–63 (noting the substantial evidence of implicit attitudes producing discriminatory behavior and discussing potential ways to alter implicit bias in decision making); Kang & Banaji, *supra* note 4, at 1064–66, (discussing how implicit bias can help us revise the meaning of affirmative action and perhaps even provide a guide to ending affirmative action).

Disparate impact and affirmative action are complementary responses to the problem of persistent discrimination. Disparate impact aims to eliminate it indirectly, while affirmative action counters its effects. The logic of this claim does not turn on whether discrimination is caused by conscious or unconscious bias. The same sort of argument could be made without any reference to unconscious bias. Evidence of covert yet conscious bias would provide as much justification for disparate impact doctrine and affirmative action as would a showing of unconscious bias. The issue is the extent of contemporary discrimination, not whether the attitudes and beliefs that cause it are conscious or unconscious.

So, as we saw in the case of Title VII and equal protection doctrine, although unconscious bias may initially seem to play some special role in antidiscrimination law, on closer examination, it raises no issues distinct from those associated with bias that is covert yet conscious.¹⁶⁵ The political legitimacy of affirmative action and disparate impact may depend on beliefs about the extent of current discrimination. But the validity of such policies cannot plausibly turn on whether that discrimination stems from bias that is unconscious versus merely concealed.

III. POLITICAL APPEAL

The ascendance of the unconscious bias discourse is a consequence of the politics of debate about racial inequality. A discourse centered on unconscious bias is more rhetorically appealing than one that highlights the continuing significance of conscious racial bias. People may more readily accept the existence of widespread unconscious bias, and even that they harbor such bias, than they would parallel claims about conscious bias. Below, we discuss four components of the appeal of the unconscious bias discourse.

A. *The Stigma of Racism*

Assertions of conscious bias are politically incendiary.¹⁶⁶ A claim of covert conscious bias impugns one's character in two distinct ways. It implies not only that one has lied, but that one has done so in order to conceal something despicable. Such implicit accusations are a necessary component of a covert

¹⁶⁵ Alexander v. Sandoval, 532 U.S. 275, 293 (2001).

¹⁶⁶ See generally FORD, *supra* note 28 (describing the controversial nature of asserting conscious racial bias against others).

bias discourse. To emphasize covert bias is to say not only that people discriminate but that they lie about doing so, to brand them both liars and racists. A charge of covert bias is thus doubly stigmatizing, a moral indictment that is not borne lightly.

Claims of pervasive unconscious bias, in contrast, do not morally indict anyone. The unconscious bias discourse presumes everyone's good faith. It is a story in which there are no villains, liars, or racists. The unconscious bias discourse depicts discrimination as a societal affliction, an illness from which we all suffer. It is not something for which one could be held morally accountable (certainly so long as one was unaware of his or her unconscious bias). The medicalization of bias makes it seem more akin to the flu, a virus run amuck in a population unaware of its existence. Indeed, this is the characterization of unconscious bias put forth by some leading legal scholars.¹⁶⁷

The unconscious bias discourse thus carefully sidesteps perhaps the most explosive landmine in contemporary racial debate: the accusation of racism.¹⁶⁸ While unconscious bias might be viewed as evidence of racism, it is the sort of racism that, paradoxically, leaves no one feeling attacked or morally indicted. One need not feel singled out as a result of one's unconscious bias if the overwhelming majority of the population, and even a substantial percentage of African Americans, have been found to harbor the same sort of anti-black bias. The unconscious bias research thus severs the link between the moral blameworthiness of the individual, and the wrongness of the resulting discrimination in the idea of "hate the sin, but love the sinner." The unconscious bias research makes it possible to do both.

B. The Denial of Discrimination

Because the unconscious bias discourse does not imply that one is a liar or a racist, there is less reason to resist the claim that implicit bias thrives somewhere within one's self. Indeed, unconscious bias discourse arguably makes it possible for people both to hear that they are racially biased and to exhibit their nonracist commitment by vowing to seek personal and societal healing. The idea, again, is that unconscious bias is a form of illness, an affliction that many may suffer, and from which all would prefer to be free.

¹⁶⁷ See, e.g., Kang, *Trojan Horses of Race*, *supra* note 4, at 1553–57 (discussing how viewers of local news "download" a type of "Trojan Horse virus" that serves to increase implicit bias).

¹⁶⁸ FORD, *supra* note 28, at 188–95.

The identification of unconscious bias thus, ironically, provides an opportunity for the affirmation of one's self-image as nonracist.

An accusation of conscious bias, in contrast, would generate a need to defend oneself, to deflect the charge of racism. The most likely response would be to deny the fact of discrimination altogether. Imagine, for example, a company whose leadership is told that the underrepresentation of women and minorities in the company's management ranks likely reflects intentional race and sex discrimination by existing managers, nearly all of whom are white men. The company would be understandably resistant to that explanation for group disparities, even if the company has already decided that it wants to increase the numbers of minority and women managers. The top-level management might not want to accept that image of the corporation. And it would be nearly impossible to sell that account to the mid-level managers who make most of the hiring and promotion decisions. Just as no individual would readily accept such a characterization of oneself, so too might corporate management be unwilling to swallow such an unfavorable depiction of its workforce. Imagine the response to a diversity consultant who, in a presentation to management, states forthrightly that the problem is that "many of you are knowingly racist and sexist." It is difficult to imagine such a consultant being hired, much less winning the confidence of firm management.

If the consultant did put forth a conscious bias explanation for race and sex disparities, one would expect a vigorous debate about the causes of minority and female underrepresentation. Corporate management would identify a whole slew of explanations other than discrimination. The company would emphasize race and sex differences in qualifications, family responsibilities, commitment to the workforce, and so forth. Individual managers would note the lack of evidence that any of their hiring or promotion decisions were discriminatory.¹⁶⁹ Such a dispute about the causes of group underrepresentation would diminish the likelihood that the problem would ever be addressed. The consultant's effort would founder at the problem-definition stage, and never reach the solution stage.

The same sort of difficulty would arise in the case of racial profiling. A researcher who asserts that the disproportionate investigation of racial minorities stems from intentional discrimination on the part of law enforcement officers is not likely to get very far with the department brass,

¹⁶⁹ Alternatively, the company might acknowledge the problem, but attribute it to a few bad apples, such as the unquestionably racist and sexist manager.

much less the patrol officers who make the stop-and-search decisions. The inevitable reaction of the department would be to deny the existence of racial profiling, and attribute the disproportionate investigation of racial minorities to other factors, such as racial differences in crime rates.

In contrast, a consultant who explains the problem in terms of unconscious bias would probably encounter less resistance. Company management and law enforcement personnel alike would be more open to the idea that discrimination persists within their organization when that discrimination is a result of unconscious bias. And if the organization can accept the fact of discrimination, it would be more likely to do something about it.

C. The Denial of Bias

Another way of avoiding the stigmatizing accusation of racism that accompanies a charge of covert, yet conscious, discrimination would be to acknowledge the fact of discrimination but to characterize it as “rational” rather than “bigoted.” This response might not have as much traction in institutional settings, where liability attaches to the fact of discrimination. But in other settings, where the question is not legal liability so much as moral indictment or political acceptability, the reasons for discrimination may matter. Discrimination is most morally objectionable when it reflects a rank form of racism, either a desire to disparage a group or the belief that the group is inherently inferior in some fundamental way.¹⁷⁰ Consider the simple case in which someone crosses a street in order to avoid a group of young black men approaching from the other direction. If the decision to cross the street reflects a more generalized or categorical aversion to black people, then the pedestrian’s decision reveals his racism. But, of course, the pedestrian would likely disavow any such despicable attitude. He would characterize the decision as a rational, self-protective measure, a sensible, if unfortunate, accommodation to the demographics of violent crime.¹⁷¹ Similarly, a cab driver who declines to pick up young black men at night would deny any racial antipathy and profess that his only desire was to avoid being a victim of

¹⁷⁰ See generally GEORGE M. FREDRICKSON, *RACISM: A SHORT HISTORY* (2002) (defining racism as a system that establishes a permanent racial hierarchy reflecting the laws of nature or decrees of God, so that stigmatized groups are unable to change their status within the dominant group).

¹⁷¹ See Jody Armour, *Race Ipsa Loquitur: Of Reasonable Racists, Intelligent Bayesians, and Involuntary Negrophobes*, 46 *STAN. L. REV.* 781, 781–83 (1994) (describing a hypothetical situation in which a woman justifies shooting a black man out of fear, claiming that her “consideration of the victim’s race was reasonable because blacks commit a disproportionate number of violent crimes”).

violence, which is disproportionately committed by young black men.¹⁷² In each of these cases there not only would be an admission of discrimination, but also an assertion that the decision was not tainted by the group animus, dislike, or irrationality that is the hallmark of racism. The discriminator took account of race, he would assert, only because race was relevant, a useful means of achieving the unquestionably legitimate goal of personal safety.

The claim that an act of discrimination was rational and, hence, nonracist raises two types of questions, one empirical and the other conceptual: First, was the discriminatory decision genuinely rational? Second, if it was, does its rationality exempt it from a characterization as racist? Neither question is easy to answer. Because young black men do commit a disproportionate amount of street crime,¹⁷³ one cannot dismiss out of hand the possibility that a coolly rational cab driver would disfavor some passengers because they are young, black, and male. Yet, the reality that race, age, and sex may be useful proxies for the likelihood of criminal wrongdoing begins rather than ends the inquiry. The cab driver might vastly overstate the usefulness of those crude indicators. Even in the unlikely event that everyone agreed that the cab driver behaved rationally, there would be disagreement about whether such rational discrimination embodies the racism to which we are all opposed.

Merely undertaking these sorts of inquiries would undermine the consensus in opposition to particular practices. Disputants would disagree about the empirics, about whether a discriminatory decision was rational. They would also disagree about whether a decision that is not wholly irrational warrants the condemnation reserved for racist practices. The fracturing of the consensus, in turn, would drain the moral force from the injunction not to discriminate. The nondiscrimination mandate would begin to seem less a categorical moral absolute, and more a big tent that covers alternative conceptions of what it might mean “not to discriminate.” Racial discrimination might come to seem more a continuum—subject to complicated empirical and moral assessment—than a discrete practice or state of mind to which we all share opposition.¹⁷⁴ This sort of process would invariably stall or thwart the social change

¹⁷² FORD, *supra* note 28, at 68.

¹⁷³ See, e.g., R. Richard Banks, *Beyond Profiling: Race, Policing, and the Drug War*, 56 STAN. L. REV. 571, 581 (2003) (noting that black and Latino adults are more likely than whites to frequently use cocaine).

¹⁷⁴ This sort of discussion would also potentially shift attention from the discriminator to the discriminatee. If the decision maker is responding to the reality of substantial group disparities, a question is likely to arise about the source of those disparities. Why do young black men commit such a disproportionate amount of street crime? Why are young black women so much more likely than white women to have a child without being married?

envisioned by those who asserted the significance of racial bias and discrimination.

An unconscious bias discourse, in contrast, neatly sidesteps these impediments to social change. Because a charge of unconscious bias does not elicit the sort of denial prompted by a claim of conscious bias, underlying disagreements about the contours of the nondiscrimination mandate or the nature of our moral opposition to racism are never flushed to the surface. If people do not vigorously contest the assertion of unconscious bias, there is no need to define the concept of bias more precisely, or to question whether a particular manifestation comes within the scope of the nondiscrimination mandate. Moreover, laypeople presume that any discrimination that results from unconscious bias is necessarily irrational; just as a computer virus may cause a computer's operating system to malfunction, so too does unconscious bias hijack otherwise rational thought processes.¹⁷⁵

D. The Historical Narrative

Finally, the unconscious bias discourse is appealing because it embodies a comforting narrative about the progress of our nation in overcoming its primal sin of slavery and racism. Everyone looks back on the period of slavery and Jim Crow as a shameful period of American history. The civil rights era represents the triumphant repudiation of that ignominious past. The big question though is how far our society has progressed from the era of de jure segregation.

The unconscious bias discourse offers a startlingly rosier account of our historical progress than would a parallel discourse of covert yet conscious bias. The ascendance of the unconscious bias discourse implies that bias lives on only beneath our awareness. In this account, covert conscious bias was vanquished long ago with the oppressive racial classifications of the Jim Crow era. Indeed, this is the interpretation explicitly embraced by a number of legal scholars and social psychologists who proclaim the importance of unconscious bias.¹⁷⁶ In their view, conscious bias has succumbed to the force of moral

¹⁷⁵ Some legal scholars make this assumption as well. See Kang, *Trojan Horses of Race*, *supra* note 4, at 1508 ("The point here is not merely that certain mental processes will execute automatically; rather it is that those implicit mental processes may draw on racial meanings that, upon conscious consideration, we would expressly disavow. It is as if some 'Trojan Horse' virus hijacked a portion of our brain.").

¹⁷⁶ See, e.g., Greenwald & Krieger, *supra* note 4 (discussing the role of implicit bias in producing discriminatory behavior and lacking discussion on the role of covert, conscious bias); Kang & Banaji, *supra* note 4 (describing the pervasiveness of implicit bias only).

norms that denounce it and legal prohibitions that prohibit it. The persistence of discrimination, in this account, reflects the development of a new form of bias that only recently has been scientifically documented.¹⁷⁷ The disease analogy figures prominently. Unconscious bias is the recently diagnosed disease that we now must mobilize our resources to fight and conquer, just as we did its cousins a generation or more ago.

A discourse of covert yet conscious bias, in contrast, would suggest that we are not as far along the road to the colorblind society envisioned by the civil rights generation. If covert yet conscious bias remains widespread, then the civil rights era ideal of colorblindness has not been as fully internalized as many would like to think. Perhaps old-fashioned racism, though rarely expressed openly in polite company, remains more of a problem than politically correct proclamations of the importance of treating people as individuals would lead us to believe. Or perhaps people discriminate not because they cling to the racism of yesteryear, but simply because they think it makes sense to do so, because it helps the cab driver and the pedestrian avoid violence, or because it helps the doctor make the best treatment decisions for his patients. In this account, bias is not something solely within one's head. It is also a response to a racially divided society. If racial discrimination is ever rational, it is so because race matters in our society, because race is intertwined by class, culture, opportunity, and well-being.

A conscious bias discourse would be more sobering, perhaps less optimistic, than a discourse centered on unconscious bias. Where the unconscious bias account depicts bias and discrimination as an unfortunate residue of earlier eras, according centrality to conscious yet covert bias renders our challenges more the object of continuing dispute and disagreement. A conscious bias discourse would imply that we have yet to agree on a goal, or more precisely that we need to work out our goals context by context, and that a meta-principle such as nondiscrimination or anti-racism cannot possibly bear the necessary moral, political, and legal demands that we place on it.

* * *

In sum, then, the ascendancy of the unconscious bias discourse reflects its rhetorical appeal. It thrives and gains traction where a conscious bias discourse would encounter resistance and denial. It allows people to acknowledge the salience of racism without having to accuse anyone of

¹⁷⁷ Kang & Banaji, *supra* note 4, at 1064.

deliberate wrongdoing. Emphasizing unconscious bias thus sustains and facilitates an ostensible consensus that racial inequality remains a problem.

Why then would we critique such a politically palatable approach?

IV. AGAINST BIAS

We critique the unconscious bias discourse because it is as likely to subvert as to further the goal of substantive racial justice. A narrow focus on the IAT may fail, as its empirical claims receive greater scrutiny, or require scholars to abandon the impartiality that is the hallmark of the scholar's commitment to truth. But even emphasizing unconscious bias more generally would be a mistake. The unconscious bias discourse prompts people to acknowledge the persistence of the race problem by misdescribing it. The unconscious bias approach not only discounts the persistence of knowing discrimination, it elides the substantive inequalities that fuel conscious and unconscious bias alike. Racial injustice, in our view, is a matter of pervasive substantive inequalities, the amelioration of which should be the goal of reform efforts.

Yet, if the unconscious bias approach succeeds, it would be because people accept it as a compelling account of racial inequality, and the eradication of bias as an appropriate goal of racial justice efforts. Once people accept the unconscious bias story, they will likely pursue its logical implications, which will not serve the cause of racial justice and may in fact undermine it.

A. *The IAT and the Risk of Failure*

1. *Empirical Uncertainty*

Emphasizing the IAT puts anti-racist law reform on a shaky empirical foundation, as a proposed reform may fail, or at least be undermined, if the empirical claims should prove unfounded. The findings of the IAT are unquestionably intriguing, but whether they demonstrate some form of unconscious bias is far from clear.

The problem that has received the least attention is the one on which we have focused. The significance of unconscious bias depends on its being distinct from conscious bias. However, because social desirability pressures cast doubt on the validity of paper-and-pencil measures of conscious bias, it is extraordinarily difficult to uncouple implicit bias from conscious yet covert bias. The IAT may measure unconscious bias, but then again it may be better

understood as a subtle measure of conscious bias that study participants are unable to conceal.

Nor does the IAT research convincingly distinguish associations that the test taker has internalized and believes from those that he simply recognizes as common associations made in the culture.¹⁷⁸ One might recognize that whiteness is associated with purity, for example, without believing that whiteness is pure. Similarly, one might recognize that in American society African Americans are associated with negative attributes, without believing, even unconsciously, that African Americans actually possess those attributes. The distinction here is not between a belief that is cultural and one that is individual.¹⁷⁹ The issue, rather, is the existence of any individual belief about the characteristics of particular social groups. This issue is important because discriminatory behavior is probably more likely to result from what one believes to be true, rather than what one recognizes others as believing to be true.

However individuals' implicit associations are characterized, the pivotal empirical question is whether they prompt discriminatory behavior. This question remains more open than often assumed.¹⁸⁰ The IAT may eventually prove a useful predictor of discriminatory behavior, but cannot now be comfortably described as such.

These ambiguities and uncertainties have been glossed over in most popular accounts of the IAT,¹⁸¹ but that's in large part because so far many commentators have worked hard to make sure the theory isn't threatening: they have insisted that it shouldn't be used to screen employees or jurors and they've actually vowed to testify against its use as evidence in legal

¹⁷⁸ See Mitchell & Tetlock, *supra* note 4, at 1049 ("It is this disjunction between conscious and unconscious attitudes that often elicits the most discomfort and disbelief among subjects, causing subjects to question the validity of IAT.").

¹⁷⁹ See Banaji et al., *supra* note 19, at 284–85 ("[O]ur basic point remains that it is less sensible to think of a sharp line between person and culture when thinking about implicit cognition.").

¹⁸⁰ Many researchers have attempted, during the past decade, to establish a link between IAT scores and discriminatory behavior, but thus far have not achieved the success for which they hoped. The most commonly cited paper that purports to substantiate the IAT as a predictor of discriminatory behavior is a meta-analysis of mostly unpublished studies, which itself remains unpublished, even as it has been cited extensively (in the legal literature in particular) for the last several years. Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity*, 96 J. PERSONALITY & SOC. PSYCHOL. (forthcoming 2009), available at <http://faculty.washington.edu/agg/pdf/UUIAT3.complete.30Dec08.pdf>.

¹⁸¹ See sources cited *supra* note 5.

controversies.¹⁸² But as the IAT makes its way into concrete policy disputes, or is advanced as a justification for controversial policy reform (e.g., regarding affirmative action or disparate impact), the research will certainly be subject to more intensive scrutiny by opponents of the policy reforms. If the more ambitious claims of the theory do not withstand critical scrutiny, then the policy reforms could fall along with them. Thus far, much of the discussion about the IAT has the feel of preaching to the choir, highlighting the pervasiveness of racism for those who already share that belief. But if unconscious bias research does anything more than tell some people what they already want to believe, it will have to convince skeptics as well.

2. *The Scholarly Role*

Linking law reform to a single measure such as the IAT also may put some scholars in a bind. Imagine, for example, an argument in favor of affirmative action that is premised on the findings of the IAT. If it turns out that the IAT is better understood as a measure of conscious psychological processes, or does not, in fact, predict discriminatory behavior, then the policy proposal would be called into question as well. If the research is inconclusive, the temptation would be nearly irresistible to shade the evidence. As a means of preserving the argument in favor of affirmative action, the scholarly advocate might understate the ambiguity of the research findings, perhaps by giving less attention to evidence that contradicts the claim on which the affirmative action proposal rests. A well-intentioned scholarly advocate might unduly discount the covert bias interpretation of the IAT, or exaggerate the evidence that the IAT predicts behavior.

Even if such partial presentation of evidence is appropriate for partisan policy advocates, it is not for scholars. Such an approach is inconsistent with the scholarly role, which comes with an obligation to present evidence with full disclosure of its nuances, uncertainties, and ambiguities.¹⁸³ Legal scholars

¹⁸² See Vedantam, *supra* note 5 (describing Mahzarin Banaji's argument against the use of the IAT as a selection tool for proving discrimination and her vow to testify in court against any attempt to use the test to identify based on individuals); Dixit, *supra* note 6 ("Yet the test's creators are extremely wary about unleashing the tool they've created. Banaji has threatened to testify in court against efforts to use her test in real-world situations.").

¹⁸³ Legal scholars should not approach a controversy as would a lawyer who is representing a client, or as would a politician or activist who seeks to further the interests of his constituency. It is admittedly difficult for legal scholars to maintain this distinction; after all, nearly all of us are trained as advocates, and many continue to represent clients. Advocates depict the facts in a manner favorable to their client's case or to the political

should be scrupulously even-handed in the treatment of empirical evidence. We need to fully communicate the ambiguity of research findings, and not allow the desire to reach a particular position to shade our characterization of the research.

B. The Costs of Success

Even an unconscious bias approach that is not narrowly focused on the IAT comes with substantial risks. To understand these risks, we begin by noting that the unconscious bias discourse facilitates consensus about the persistence of the race problem by, ironically, misdescribing the nature of that problem. While we do not doubt the existence of unconscious bias, we do doubt that racial bias explains all or even most of the racial injustices that plague our society. Thus, the unconscious bias discourse should not be understood as representing a goal of racial reform so much as embodying a strategy. We suspect that this strategic sensibility underlies the embrace of the unconscious bias idea by many legal scholars, who hope to use the politically appealing notion of unconscious bias to leverage support for policies that would further the goal of substantive equality.

But the strategic invocation of unconscious bias brings with it substantial risks. The worry here is not failure but success. If the unconscious approach succeeds, it would do so because many people have accepted its depiction of racial injustice, having, in essence, mistaken a strategy for the goal of reform. Once that happens, true believers would likely embrace the logical implications of the unconscious bias idea, which may not serve the cause of racial justice.

1. Goal Distortion: Confusion of Means and Ends

Even if understood as strategy, uncovering and eliminating unconscious bias might, over time, come to be mistaken for the goal of racial reform. Today's means could become tomorrow's end. Consider the past century of anti-racist activism. The idea that the core of anti-racist legal and policy intervention should be the elimination of "bias" in individual cases—as opposed to a more measurable substantive goal such as integration, reparations, or equal outcomes—is itself the result of controversial decisions at earlier stages of the civil rights struggle. It is unlikely that racial justice

outcome they support; scholars should not. Similarly, politicians might describe the facts selectively, but scholars should not.

activists during the early twentieth century envisioned their goal as the elimination of bias in individual cases. Indeed, they were trying to dismantle the racially oppressive system that developed in the aftermath of chattel slavery. The idea of eliminating discrimination was a means of realizing their goal; it was not the goal itself.

Having committed to the antidiscrimination approach because it was thought to be an effective strategy, and having succeeded in convincing a large swath of the American public, we now find ourselves trapped: many decisions and practices that adversely affect racial minorities do not fit neatly within the conventional antidiscrimination framework—and so, given the constraints of legal and popular discourse, we are forced to press ever more expansive, rarified, and questionable definitions of “discrimination” and “bias” in order to cover the elusive decisions and practices. Many of us even believe that the goal of anti-racism should be the elimination of individual bias. Thus, do the strategic innovations of one era become the substantive constraints of the next? The same invocation of antidiscrimination that helped topple Jim Crow now limits further efforts to ameliorate troubling racial disparities. Those judges, commentators, and citizens who believe that the nondiscrimination mandate precludes affirmative action, for instance, are not perverting the civil rights era quest for racial justice, so much as they are taking it to heart. They are accepting as substantive mandate what to a prior generation’s civil rights lawyers was a strategic innovation.

This confusion of means and ends is a natural outgrowth of the use of an argument for strategic reasons. The idea of using a theory tactically is to convince people who would not accept one’s underlying goals if they were stated forthrightly. But if the tactic works, it can work only because you have, in some sense, deceived people about the underlying goals and leveraged something else in order to secure support for them. There is no way, consistent with this tactic, to also make sure that people don’t take the theory seriously—and in the wrong direction. To avoid such a risk, one would have to admit that the real goal is something other than what it seems to be, which would, of course, undermine the persuasive effect of the rhetoric. As those of us who subscribe to the strategy are followed (and, in the course of time, replaced) by a new generation, they will, if we’ve done our rhetorical job well, sincerely believe in the unconscious bias account as a description of the problem of racial injustice.

Not only may the strategy convince others, over time we may accept it as true as well. If unconscious bias discourse yields any short-term benefits, we will, as a practical matter, be hard pressed to abandon its implications later, and, worst yet, we will most likely be too psychologically and intellectually invested in it to examine the decision critically. Even if we begin arguing for unconscious bias theory as a matter of expediency, we will likely wind up sincerely accepting it (and its limitations). The tactical theory will come to define and limit the scope of our political imagination.

As every political dictator who has required repetitive oaths of allegiance knows, if one is forced to say something often enough, one begins to believe it. Arguments first advanced for instrumental, purely strategic reasons have a way of becoming genuine commitments the longer and more ingeniously one argues for them. Anyone with knowledge of the practice of law recognizes this phenomenon: a lawyer begins using an argument instrumentally, saying whatever is likely to convince a judge or jury to find in favor of her client. She makes seemingly impassioned arguments for clients whose cases she recognizes as weak. It's her job—that's it. But over time, slowly but surely, as the controversy drags on she comes to believe in the rightness of her client's position, in the force of claims that she would have earlier regarded as fanciful. This same phenomenon is likely to occur with respect to claims made by legal scholars for strategic reasons.

Theories build upon and develop prior theories. If we start down the wrong theoretical path in the hope of tactical gains, we may well find ourselves without the intellectual sense of direction necessary to change course later. Indeed, we worry that this has already happened. When law professor Charles Lawrence articulated the idea of unconscious bias more than twenty years ago, he did so in order to argue for a disparate impact standard in equal protection jurisprudence after the Court rejected such a standard in *Washington v. Davis*.¹⁸⁴ It was clear that Lawrence favored the disparate impact standard for consequentialist reasons and would have favored it regardless of his beliefs about unconscious bias.¹⁸⁵ Indeed, much of his defense of disparate impact focused on the difficulty of proving consciously held discriminatory intent.¹⁸⁶ Now, a generation later, many of the scholars following in Lawrence's wake

¹⁸⁴ See generally Lawrence, *supra* note 3 (rejecting the doctrine of discriminatory purpose established in *Washington v. Davis*, 426 U.S. 229 (1976), and instead stressing the importance of the equal protection doctrine coming to grips with unconscious racism).

¹⁸⁵ *Id.*

¹⁸⁶ *Id.*

have advanced much more empirically presumptuous, individually probing, and cognitively focused versions of the unconscious bias idea. No doubt they would say they have refined and improved Lawrence's pioneering theory, but we would argue that they have made it worse. They have magnified the significance of the mental state of governmental decision makers, even as they have obscured the underlying goal of substantive equality that animated Lawrence's approach.

Our claim that strategic reliance on the unconscious bias discourse would likely transmute into an end in itself raises a pivotal question: what would be so bad about that? Why not conceive of racial justice as the eradication of unconscious bias?

2. *Undesirable Outgrowths*

If people buy into the idea that the problem of racial injustice is fundamentally a problem of unconscious bias (which they must if the strategy is to be successful), then they are likely to elaborate aspects or implications of the theory whose links to racial justice are more and more attenuated, and which may be undesirable in their own right. Some true believers may embrace the view of bias as a quasi-medical ailment, which we term the medicalization of racism. Others may advocate a narrow focus on instrumentally rational decision making, which we term "technocratic authoritarianism." Perhaps the majority will focus heavily on mental state and waste time in intractable conflicts over peripheral issues. In none of these scenarios would the cause of racial justice be well-served.

One risk of unconscious bias discourse is that it will further a trend toward technocratic authoritarianism. Most accounts of unconscious racial bias define it as one type of a broader category of "bias." There are IAT tests for implicit biases against women, religious minorities, the disabled, the overweight, the elderly, and even Republicans and Democrats. Consistent with the multiplicity of types of bias identified by the test, some scholars seek to combat "bias" generally in decision making.¹⁸⁷ The implication is that after attacking racial bias, we should try to enforce a norm of technically rational decision making more generally.¹⁸⁸ Subsequent reform efforts will aim to eradicate any

¹⁸⁷ See generally Kang, *Trojan Horses of Race*, *supra* note 4 (discussing the real-world consequences of implicit biases).

¹⁸⁸ *Id.*

decision-making process that is not rational in a narrow instrumental sense.¹⁸⁹ This is a logical, if not inescapable, implication of the metaphor of “bias,” which is meaningless without some notion of the straight and true. When one cuts fabric on the “bias,” one cuts on a diagonal to the weave. But what is the “weave” of proper decision making? The most plausible answer entails some form of instrumental, means/end rationality. The goal of eliminating bias then becomes an effort to enforce a principle of strict rationality, the sort of instrumental logic associated with computers and technology.

Technocratic logic, however, may well run counter to the goals of anti-racism on many substantive questions. For example, while anti-racism may oppose racial profiling, technocratic authoritarianism may support it if, for example, differences in rates of criminality across groups render the consideration of race a rational means of efficiently apprehending wrongdoers. Similarly, whereas anti-racism may support affirmative action, technocratic authoritarianism may question the rationality of allocating scarce jobs or positions in selective schools to people whom objective criteria accurately predict will perform less well than some of those who are turned away. Technocratic rationality may condemn affirmative action as an irrational practice, driven by ideological dogmatism, neurotic guilt, and a romantic egalitarianism that the facts simply do not support. Conversely, whereas anti-racists, in the name of autonomy and pluralism, may be willing to allow businesses to pursue idiosyncratic personnel policies provided they do not further racial injustice, technocrats inspired to eliminate bias may well wish to impose “best practices,” defined in empirical and technocratic terms in order to ensure that valuable social resources are not wasted by technically capricious or incompetent managers.

We suspect that much of the support for unconscious bias theory reflects anti-racism that is understood in terms of technocratic logic. We do not think anti-racism should be so tightly bound up with technocratic rationality. Following social theorists from Max Weber to Michel Foucault, we worry about the alienation and disenchantment that follows from the technical bureaucratization of social life. At least in the case of private decision makers such as employers, there are good libertarian and humanistic reasons to permit and to value intuition, aesthetic judgment and even whimsy—none of which could be defended in technocratic terms. Insisting that employers defend decisions that adversely affect members of historically subordinated groups, as

¹⁸⁹ *Id.*

we have argued Title VII doctrine does, is a relatively mild restriction on decision-making prerogatives, justified in terms of combating widespread social animus and promoting the integration of isolated and relatively impoverished groups. But expanding this limited injunction to combat “bias” generally would be a major intervention in favor of technocratic and centralized control of the market economy and autonomous institutions. Any such intervention should be justified in its own right—and should respond to the many and varied criticisms of bureaucratic logic, technological fetishism, and centralized control of markets. It should not be permitted to ride the coattails of anti-racism.

A second worry is that unconscious bias theory makes racism seem more a medical problem than a social problem.¹⁹⁰ None of the psychological researchers explicitly characterize unconscious bias in these terms, but as the idea of unconscious bias has made its way into legal and popular discourse, the notion has been shaped by the metaphor of illness. One legal scholar has explicitly compared unconscious bias to a computer virus.¹⁹¹ Even if scholars explicitly disclaim any medical characterization of unconscious bias, others may not. The idea of illness is too natural a fit. The notion of a bias that resides within and distorts one’s thinking and feeling practically cries out to be viewed as a medical problem.

The tendency toward medicalization is consistent with the fetishism of science and technology that’s the zeitgeist of our era—we tend to think that almost everything is reducible to mathematics and therefore to technological fixes. If racism is a psychological illness, a pharmacological cure can’t be far behind: *if you’re feeling especially intolerant or if bigotry or xenophobia is interfering with your job performance or personal relationships, ask your doctor about a little rainbow-colored capsule available by prescription.* More likely though, the remedy will take the form of counseling or therapy. As a cure for racism, psychotherapy is no less worrisome than a pharmacological fix.

Medicalization, whether in the form of counseling or pharmacology, is an especially insidious form of social power and coercion. Modern bureaucracies

¹⁹⁰ There is already evidence of this approach. See, e.g., Alvin F. Poussaint, *Is Extreme Racism a Mental Illness?*, 176 W. J. MED. 1 (2002) (“It is time for the American Psychiatric Association to designate extreme racism as a mental health problem by recognizing it as a delusional psychotic symptom.”).

¹⁹¹ See generally Kang, *Trojan Horses of Race*, *supra* note 4.

use techniques of knowledge production to exercise power over individuals, all the while denying that power is in fact being exercised at all. Compare an edict that all people with unacceptable views be reeducated in mass camps and a medical diagnosis that people who express unacceptable views and impulses suffer from an ailment and should accept “help” in the form of therapy sessions in rehabilitation clinics. Of course, the medical diagnosis is not a formal edict, but it can be given teeth if we think that “sick” people should not be trusted to exercise authority or interact with the public. It should not comfort us that we agree that the censored views really are repugnant; of course the censored views will be extremely unpopular and reprehensible, at least at first. But once the precedent has been established—that we should police thoughts for implicit “biases” and then “help” people to overcome them—the expansion of this form of social control will be difficult to halt. The beginnings of this process are already apparent.

Already the science of implicit bias is used as part of mandatory sensitivity training sessions in many workplaces. This is a soft form of coercion: no one is told that they are evil or culpable—instead they are told that their unconscious mind may be working against their better selves. Anti-racism does not require this sly and subtle but also extremely invasive form of social control: we can and should regulate objective decisions and substantive outcomes, but we should not seek to monitor and control attitudes and thoughts.

Medicalization is also worrisome because it inclines people to disclaim responsibility for their own decisions and actions. Consider the number of people who have, no doubt on the advice of public relations consultants, vowed to enter sensitivity training or therapy in order to “work through” their prejudices after being overheard making some blatantly racist statement. Mel Gibson spews anti-Semitic bile at police officers who arrest him for drunk driving and later, in a fraught confession, blames alcohol (as if liquor can somehow spontaneously generate beliefs that were not held before) and announces his intention to seek treatment.¹⁹² Michael Richards abuses black audience members with racist invective and later announces plans to enter therapy to work through his personal issues.¹⁹³ The appeal of the medicalization excuse fits neatly with our critique that the IAT may be

¹⁹² Denise Mann, *Is Alcohol a Truth Serum?*, CBS NEWS, Aug. 3, 2006, <http://www.cbsnews.com/stories/2006/08/03/health/webmd/main1864620.shtml>; Buck Wolf, *Gibson Seeks Rehab*, ABC NEWS, July 31, 2006, <http://abcnews.go.com/entertainment/Story?id=2257775&page=1>.

¹⁹³ *Richards Still Sorry, Goes into Therapy*, USA TODAY, Nov. 27, 2006, at 3Q.

measuring attitudes or beliefs of which people are perfectly well-aware. In both cases, it becomes easier for people to evade responsibility for their own behavior. We favor more accountability rather than less.

On the flip side of the medicalization of bias, we worry about bias becoming de-stigmatized. The illness metaphor alleviates responsibility. It could also contribute to bias being viewed as ineradicable. The fact that most people are found to have some form of implicit bias could bolster the notion that implicit bias is not so bad after all. Such a development would be undesirable as well.

The focus of the unconscious bias discourse on the hearts and minds of putative perpetrators is part of an especially aggressive regulation of the self that is the dark side of pop psychotherapy's wealth of twelve-step programs and self-help treatises. It's not enough that people avoid perpetuating racial injustice. Now the goal is to purge the unconscious mind of the body politic of unacceptable biases and prejudices. And "unacceptable" comes pretty quickly to mean anything that can't be justified in scientific, economic, rational, numerical terms. It's a mandatory technocracy, with antidiscrimination law conscripted into the thought police. Such is not our vision of racial justice.

Finally, the most worrisome aspect of the unconscious bias discourse is that it is likely to reinforce a misguided preoccupation with individual acts of discrimination. The effort to root out individual bias and eliminate discrimination is so much a part of how we think about racial inequality that it may seem odd to criticize it. But we wonder whether the time may have come to abandon the antidiscrimination framework that the unconscious bias discourse exemplifies. There are numerous problems with the antidiscrimination approach.

First, as we have already discussed, individual acts of discrimination cannot plausibly explain the persistence of our society's most troubling racial disparities. In this sense, the goal of eliminating discrimination is too modest, not ambitious enough. If we are legitimately concerned about substantive disparities, then we should not content ourselves with only alleviating those disparities that are traceable to individual bias.

Second, the unconscious bias discourse siphons energy away from substantive reform projects by inviting an intractable, and time consuming, inquiry into the inscrutabilities of mental state. The effort to identify prohibited mental states is a fool's errand fraught with both practical and

conceptual difficulties. The practical difficulty is that proving the prohibited state of mind is extraordinarily difficult. We simply don't have access to what people are thinking, much less the unconscious recesses of their minds, and the attempt to plumb their mental states is a daunting task with little payoff. Even if we could ascertain precisely the contents of a person's mind, such perfect knowledge would only highlight the conceptual indeterminacy of the idea of discrimination on the basis of race. The state of mind that is prohibited is subject to disagreement in part because the notion of race is itself subject to disagreement. Culture, language, class, appearance—all these potential bases of discrimination might, or might not, support a claim of racial discrimination. There is simply no objective way of determining what counts as discrimination on the basis of race.

Third, the unconscious bias discourse may further legitimize an antidiscrimination framework that constrains desirable policy initiatives. Many efforts to promote greater substantive racial equality will, almost by definition, take account of race. The antidiscrimination framework unjustifiably casts such efforts as suspect. Seizing on the possibility of unconscious bias reinforces a framework that we should ultimately be attempting to displace.

CONCLUSION

We suspect that the goals underlying the growing focus on unconscious bias are laudable: troubling racial inequalities persist and existing legal doctrine and policy initiatives have not delivered us to the promised land of racial justice. The unconscious bias discourse is a politically palatable way of reminding people of the continuing salience of race in American society. But it spurs people to acknowledge our continuing race problem by misdescribing it. Today's racial injustice is largely a matter of substantive racial inequalities—not the hidden bad attitudes of individuals. Even if unconscious bias is as prevalent as its theorists claim, it does not plausibly account for the persistence of such disparities.

Instead of pursuing ever more obscure and elusive forms of bias and discrimination, we should make substantive arguments in defense of substantive policy goals—arguments that must stand or fall on their own.

Because it misdescribes the problem of racial injustice, unconscious bias theory inspires misguided reform efforts. It fuels fruitless attempts to ferret out individual bias and places too much emphasis on individual acts of discrimination. Ultimately, the unconscious bias approach may obscure, or even undermine, the substantive goals of racial justice.