AGAINST PREJUDICE

Stephen M. Rich
University of Southern California
Gould School of Law
srich@law.usc.edu

INTRODUCTION

Contemporary cognitive psychologists propose a transformation in our understanding of prejudice. No longer are we directed to view prejudice as "thinking ill of others without sufficient warrant,"[1] or to equate it with conscious antipathy toward members of discrete social groups for which the agent lacks a legitimate reason. Today psychologists define *prejudice* more broadly as a negative or ambivalent attitude, attributable to ordinary and even automatic mental processes, that is capable of influencing cognition and behavior beyond the agent's conscious awareness or control.[2] What the law ought to do with this "new prejudice" is already a matter of great controversy among psychologists and legal scholars alike. Some have argued, I think rightly, that the new account of prejudice directs us to define *discrimination* to include conduct motivated by unconscious attitudes and stereotypes. Others have countered that to do so would undermine our understanding of why discrimination is wrongful, threatening to conflate culpable thoughts with culpable deeds. Many interesting and important questions issue from this debate. This Essay will address one that has as yet received no direct attention, which is whether the new account of prejudice should give us cause to reject prejudice as the core and defining concern of antidiscrimination law. I argue that it does present us with just such an opportunity and that this is an opportunity we should take.

---

[1] Gordon W. Allport, The Nature of Prejudice, 7 (1954).

[2] Following the contemporary psychological literature, I mean by *prejudice* a negative or ambivalent belief or attitude, whether consciously (*explicit* prejudice) or unconsciously (*implicit* prejudice) held, directed at individual persons or groups of persons on the basis of their group status. As an attitude, prejudice is often associated with animus or hostility, though it may include lesser aversive attitudes. But I do mean to distinguish between prejudices *against* (properly, *prejudices*) and prejudices *for* (*preferences*). In defining prejudice to include beliefs (rather than limiting it to attitudes), I have departed from the convention in the psychological literature that defines prejudices as attitudes and stereotypes as beliefs; I do this because we often speak of beliefs as prejudices. Prejudices will often reflect *stereotypes*, by which I mean beliefs, whether consciously (*explicit* stereotypes) or unconsciously (*implicit* stereotypes) held, associated with persons or groups of persons on the basis of their group status, which may be inaccurate or, at least, overestimated when applied to individuals, and purport either to describe traits held by group members (*descriptive* stereotypes) or to prescribe behavior, traits or other qualities appropriate to them (*prescriptive* stereotypes).

I anticipate resistance. One may hold, for example, that the relationship between prejudice and discrimination is constitutive, as if prejudice were a kind of reason for acting by which conduct, though it might otherwise have been consistent with other innocent explanations, becomes deserving of the label *discrimination*. For this position the converse, of course, must also be true – that the same conduct is not discrimination if it is not motivated by prejudice. Nowhere does this objection gain more strength than in consideration of *intentional discrimination*, a phrase that the Supreme Court has used as at times synonymous with and at others a defining feature of the statutory claim of disparate treatment workplace discrimination.[3] Antidiscrimination law has identified other forms of discrimination, such as disparate impact discrimination and failure to accommodate disability or religious practice. Most students of antidiscrimination law would agree that at least one, if not both, of these forms discrimination transcends anything resembling what psychologists mean by *prejudice*. But the concept appears to hold a special place in disparate treatment doctrine because the requirement of intentional discrimination appears to endorse a cognitive interpretation of discriminatory treatment. This is a misreading of the doctrine, one that has driven legal scholarship to appeal to cognitive psychology in order to provide what ought to have proved an unnecessary correction.

To sustain a claim of disparate treatment discrimination under Title VII of the 1964 Civil Rights Act,[4] a plaintiff must prove that the defendant intentionally discriminated by treating her differently than other similarly situated persons because of her race, color, religion, sex, or national origin, and this disparate treatment resulted in an adverse employment action.[5] Sometimes prejudice matters in this determination, and sometimes it does not, depending on whether the facts of a case afford the plaintiff alternative means to prove that the defendant subjected her to disparate treatment because of her status. Its relationship to the claim is merely instrumental, not constitutive. This is no mistake. For no matter how broadly we define prejudice, and even if we include its unconscious forms, any construction of its role as a constituent element of discrimination would excuse forms of unequal treatment otherwise committed because of the victim's protected class status.

I am speaking now not only of cases in which the defendant held an ambivalent motivation for considering the plaintiff's protected class status, where it may be difficult to reach agreement regarding whether the motivation is prejudicial, but also cases in which the defendant acted out of a cynical motivation to feign compliance with the mandates of antidiscrimination law, indifferent to any harm this ultimately caused to the plaintiff, and cases in which the defendant sincerely intended to promote diversity but failed miserably and thereby injured the plaintiff. No matter how broad, any definition of prejudice will always reach its limit case, and the question for antidiscrimination law is whether to protect plaintiffs injured by conduct beyond that limit, beyond prejudice.

---

[3] Compare St. Mary's Honor Ctr. v. Hicks with Int'l Bhd of Teamsters v. US n.15.

[4]

[5] See infra Part II for a more detailed discussion of the doctrine.

In this Essay, I argue that intentional discrimination does not imply that the agent acted with prejudice. Rather, the concept of intentionality found in disparate treatment doctrine is sufficiently broad to include any deliberate act of unequal treatment of which the protected class status of the plaintiff can be proved to have caused, regardless whether the employer's motivation were conscious or unconscious, prejudicial or benevolent. Indeed, the employer's motivations are relevant only to the extent that they are necessary to establish that the nature of her disparate treatment is that she was subjected to it because of her protected class status. If she can otherwise demonstrate a causal relationship between her status and the unequal treatment, whatever innocent motivation the employer may offer should be ignored as an unwelcome distraction. The doctrine should, therefore, retain the requirement of intentionality and reject any effort to infer a requirement of prejudice.

Few have suggested agreement with this proposal.[6] Much more prevalent is the converse: a proposal to scrap intentionality and emphasize an enlarged understanding of prejudice that is informed primarily by cognitive science.[7] By suggesting that we turn to cognitive psychology to retool antidiscrimination law, legal scholars have not abandoned the concept of prejudice so much as tinkered with its meaning, enlarging it to include features such as unconscious association and attitudinal ambivalence. However, contemporary psychological descriptions of the dynamics of discriminatory behavior sit at odds with the traditional animus-based normative outlook that depicts *discriminators* as consciously motivated by hateful attitudes and irrational beliefs and *discrimination* as a social practice aimed to perpetuate group disadvantage. It is, therefore, somewhat surprising that recent scholarship explicitly puts science before normative theory,[8] claiming that a new descriptive account of discrimination must precede any revised normative account of antidiscrimination law – that is, except where they have collapsed these two pursuits entirely.[9]

While cognitive science may provide us with useful if sometimes unsettling information about the motivational dynamics of discrimination, it does not tell us why the law ought to label particular behaviors discriminatory.[10] For this reason, a few scholars have begun to call for a more robust revision to the normative framework in which we situate antidiscrimination law,[11] and a side debate has already begun about the suitability of a such proposed revision.[12] I aim to contribute to this conversation by proposing that we restrict the normative function that we assign to prejudice from one in

---

6 Strauss. Selmi.

7 See infra.

8 Krieger.

9 Kang. Greenwald & Krieger.

10 It does not even tell us that we may not differentiate between forms of discrimination on the basis that they emanate from conscious or unconscious motivations. Kang & Banaji.

11 Bagenstos. Blasi. Green.

12 Structuralism. Sturm; Bagenstos; Green.

which prejudice is viewed as a constituent element of discrimination that rationalizes and legitimates antidiscrimination law to one in which it performs a more pragmatic function in assisting us to prove certain types of claims but does not fundamentally limit the various types of disparate treatment claims that a plaintiff may bring. My argument advises caution against permitting scientific claims to supplant our understanding of the normative purposes fulfilled by the law.

Antidiscrimination law requires us to make normative choices regarding what sorts of conduct we ought to prosecute as wrongful discrimination, what sorts we may excuse, and why. The new cognitive account of prejudice will not make these choices for us, just as the traditional animus-based view has not fully constrained our understanding of what constitutes unlawful unequal treatment. Under existing legal frameworks, the concept of prejudice performs both a pragmatic and a normative function. First, pragmatically, prejudice aids litigants to demonstrate that their protected class status is the cause of their unequal treatment. In this sense, attributions of prejudice are surely useful for judging whether an agent's behaviors are discriminatory, but they are not necessary. Where a plaintiff has otherwise shown that her status dictated her treatment, she has proved discrimination whether or not she has proved prejudice. Furthermore, recent psychological research suggests that, if we took prejudice to be a necessary condition for judgments of discrimination, we may fail to judge conduct discriminatory if evidence of prejudice is inconclusive even though, without a prejudice condition, we would hold the same conduct to constitute discrimination.[13] This should raise caution that if we rely too heavily on prejudice to prove discrimination it may cease to perform its pragmatic function by frustrating the factfinding process.

Second, the concept of prejudice performs certain normative work in antidiscrimination law for courts and legal scholars seeking to rationalize perspectives on legal liability and effective remedies. Normatively, prejudice suggests a reason to hold persons morally culpable for discriminatory acts. It also situates antidiscrimination law sociohistorically, permitting us to explain it as a response to historical forms of status hierarchy and disadvantage actively perpetuated by discriminatory practices. Robert Post argues that "[a]ntidiscrimination law seeks to neutralize widespread forms of prejudice that pervasively disadvantage persons based upon inaccurate judgments about their worth and capacities."[14] The notion that antidiscrimination law prohibits unequal treatment because of prejudice crystallizes one of the founding accounts of antidiscrimination law, but this account hardly provides us with a complete picture. The "forms of prejudice" referenced by Post transcend purely cognitive descriptions of prejudice. Post himself

---

[13] See, e.g., Janet K. Swim et al., The Role of Intent and Harm in Judgments of Prejudice and Discrimination, 84 J. Personality & Soc. Psych. 944 (2003) (demonstrating that, where evidence of an agent's intentions are inconclusive, lay persons more frequently conclude that behaviors reflecting differential treatment are discriminatory than that the agent was motivated by prejudice, because they are less willing to render a negative judgment of the agent's character than to diagnose differential treatment as discrimination).

[14] Robert Post, Prejudicial Appearances: The Logic of American Antidiscrimination Law, 88 Cal. L. Rev. 1, 8 (2000).

observes that, even if one includes social stereotypes in the definition of prejudice, we must admit under a "sociological account" of antidiscrimination law that it "does not ask whether 'stereotypic impressions' can be eliminated *tout court*, but rather how the law alters and modifies such impressions."[15] For example, employment discrimination law rejects the notion that traditional gender roles should be salient in assigning work opportunities to women[16] but permits employers to invoke gender roles that reiterate stereotypic gendered assumptions in requiring men and women to conform to differing standards of proper workplace appearance.[17] Believing that a concern to prohibit prejudice alone dictates the outcome in work assignment cases may blind us to the incompetence of the same prohibition to dictate the outcome of the grooming cases. Post's observations therefore force us to reconsider the sometimes facile way in which we announce the normative underpinnings of antidiscrimination law.

The concept of prejudice does not commit us unequivocally to any particular set of normative choices; it merely provides us with a language with which to explain our normative choices. Enlarging our understanding of prejudice, while perhaps necessary as a scientific and epistemological matter, does nothing to resolve the normative tensions that have arisen from the traditional understanding. Rather, it simply generates new questions. On the one hand, to say that antidiscrimination law targets prejudice is not to say that it targets prejudice *tout court*; nor is it to say that it targets prejudice as state or mind but not as social practice. On the other hand, if in saying that antidiscrimination law targets prejudice we mean that it excuses from legal challenge differential treatment not motivated by prejudice, this is simply an inaccurate account of the scope of antidiscrimination protection and the appeal to prejudice explains nothing. Worse, it threatens retrenchment rather than expansion of antidiscrimination protections by restricting the manner in which the law constructs wrongful discrimination. For the proposed enlargement of our understanding of prejudice is itself no enlargement at all if it requires that we define discrimination in terms of its relation to prejudice, and then only through a cognitive framework.

Granted, there are dangers associated with this argument. Any rejection of prejudice as a defining element of discrimination and thereby a founding concern of antidiscrimination law risks facing the critique of "color blindness."[18] In other words, it risks being labeled ahistorical and inhumane in failing to identify a moral position from which antidiscrimination law may fulfill its function to disestablish forms of status hierarchy and group-based material disadvantage. So I must be very clear.

The argument against prejudice is not an argument for rejecting the historical framework by which we understand antidiscrimination law as an effort to alleviate social stratification by regulating the practices that create and sustain it. It is not

---

15 *Id.* at 31.

16 See, e.g., Price Waterhouse

17 See, e.g., grooming cases

18 Siegel (twice).

an argument against the view that we should understand discrimination as a particular problem for certain historically disadvantaged groups and antidiscrimination law as particularly interested in remedying these historical forms of discrimination. It is not, in other words, an argument against resting antidiscrimination law on a principle of antisubordination.[19] Rather, consistent with the antisubordination principle, the argument against prejudice rejects the notion that discrimination should be defined by a particular state of mind, specifically an agent's conscious animus toward members of certain social groups. But, unlike the antisubordination principle, the argument against prejudice does not rely on effects-based liability tests, such as disparate impact theory, to fulfill its account of discrimination.[20] I uphold the value of disparate impact theory for a full and robust approach to remedying discrimination, but it is not my concern here. Instead, my argument is that disparate treatment discrimination should not be determined by whether the agent acted out of hostility toward the target, consciously or unconsciously, but simply by examining whether the plaintiff has been treated differently from other persons because of her social group status.[21]

Ultimately, the argument against prejudice is a rejection of the cognitive turn taken in antidiscrimination law insofar as it threatens to curtail the law's normative commitments to equal treatment and antisubordination by controlling how we theorize about discrimination, including intentional discrimination. A proper accounting of the cognitive understanding of prejudice may be indispensable to the just adjudication of individual claims of disparate treatment. But the argument against prejudice rejects the notion that any cognitive understanding of prejudice should be so central to legal doctrine that it cabins what we may identify as discrimination, including disparate treatment discrimination.

This Essay will proceed in steps. [INSERT ROADMAP]

---

[19] Siegel. Balkin.

[20]

[21] In this sense, I see a strong parallel between my thesis and argument developed by Vicki Schultz in her seminal *Reconceptualizing Sexual Harassment* that hostile work environment sexual harassment should not be defined by whether the perpetrator was motivated by sexual desire, but whether perpetrator targeted the plaintiff for harassment because of her sex. Reconceptualizing Sexual Harassment, 107 Yale L. J. 1683 (1998).

I. The New Prejudice and the Cognitive Turn in Antidiscrimination Law

A. *The new prejudice*

The "new prejudice" is known by many names.[22] Most commonly, psychologists and interested legal scholars now refer to it as *implicit bias*, but this is something of an inexact phrase. The most current and authoritative sources use this phrase to refer to *implicit attitudes*, of which unconscious prejudices form only a subset, and *implicit stereotypes*. For example, in their introductory piece to the *California Law Review* "Symposium on Behavioral Realism," Anthony Greenwald and Linda Hamilton Krieger define "implicit biases" as "discriminatory biases based on implicit attitudes or implicit stereotypes" capable of "produc[ing] behavior that diverges from a person's avowed or endorsed beliefs or principles."[23] Greenwald and Krieger further explain that implicit biases are *implicit* in the "technical" sense that they operate "outside conscious attentional focus."[24] That is, they are unconscious and therefore may not be subject to the agent's conscious retrieval or awareness.

Equally important, implicit biases are *implicit* based on the testing methods by which they are observed. Rather than explicit measures that assess bias by requesting respondent to engage in self-reporting, implicit measures rely upon indirect, non-testimonial measures such as reaction times in response to priming procedures[25] or in the performance of categorization tasks.[26] Implicit measures, such as the Implicit Association Test ("IAT") are designed to observe unconscious attitudinal and stereotypic associations "in a manner not discerned by respondents."[27] Fundamental to the conclusions of implicit bias researchers, and especially relevant their research into antidiscrimination law, is the claim that implicit bias may be dissociated from self-reporting and that one should take this as evidence of the reliability of implicit measures because respondents may undermine the accuracy of self-reporting measures either by conscious deception or by falsely testifying as to the truth of their attitudes and beliefs.

---

22 Implicit prejudice. Unconscious racism. Aversive racism. Modern racism.

23 Anthony G. Greenwald & Linda Hamilton Krieger, Implicit Bias: Scientific Foundations, 94 Cal. L. Rev. 945, 951 (2006). Notably, each article published in connection with the symposium is devoted to the subject of implicit bias. [confirm Blasi & Jost]

24 *Id.* at 947.

25 *See, e.g.*, Russell H. Fazio et al., Variability in Automatic Activation as an Unobtrusive Measure of Racial Attitudes: A Bona Fide Pipeline?, 69 J. Personality & Soc. Psych. 1013 (1995).

26 *See, e.g.*, Anthony G. Greenwald et al., Measuring Individual Differences in Implicit Cognition: The Implicit Association Test, 74 J. Personality & Soc. Psych. 1464 (1998). *See also* Anthony G. Greenwald & Mahzarin R. Banaji, Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes, 102 Psych. Rev. 4 (1995).

27 Greenwald & Krieger, supra note __, at 952.

Psychological research on prejudice has a long and evolving history that appears to have been driven in large part by the need to reconcile public opinion polls showing a decline in overt racial prejudice beginning after World War II and persistent signs of "subtle" discrimination that have continued to structure relations between racial groups. Though in its infancy prejudice studies constructed prejudice as "psychopathology" and a "dangerous aberration in normal thinking,"[28] Gordon Allport's early contributions to the field set the stage for many of the developments that now seem to have crystallized with implicit bias research. Allport emphasized the normality of prejudice as a form of prejudgment, reflecting a normal cognitive tendency to categorize objects in or social field and to permit associations with the category to be transferred to examples within the category.[29] But Allport retained and expanded upon a view of prejudice as "antipathy based upon faulty and inflexible generalization."[30] Simply put, setting aside Allport's view that prejudice develops from a normal cognitive process, the core of his definition retained features of antipathy and irrationality/inaccuracy that are very consistent with common intuitions regarding why the law ought to hold someone accountable who harms another when motivated by prejudice.

The next wave of prejudice studies emphasized Allport's more radical point – that "prejudice is rooted in normal rather than abnormal processes."[31] Prejudice in some form came to appear not only as a normal outcome of cognition but, under some descriptions, as an inevitable one. During this period, two influential approaches developed. First, John McConahay and colleagues developed the Modern Racism Scale, which purports to measure covert racial prejudice by measuring the subject's attitudes toward political ideologies and assumptions reflecting support or hostility toward a progressive, egalitarian racial agenda.[32] The Modern Racism Scale was developed on the premise that, as political tides changed, people no longer feel comfortable expressing overt prejudice but are more comfortable expressing views that they purport to derive from the observation of empirical facts.[33] Therefore, modern racism is characterized by the subject's success in rationalizing and concealing by unwittingly altering the expression of racial prejudice, motivated in part to avoid attributions of bigotry.

---

28 John F. Dovidio, On the Nature of Contemporary Prejudice: The Third Wave, 57 J. Soc. Issues 829, 830 (2001). Dovidio refers to this early period as the "first wave" of prejudice scholarship in the psychology literature.

29 Allport, supra, at 20-24. See also Susan T. Fiske, Social Cognition and the Normality of Prejudgment, collected in On the Nature of Prejudice: Fifty Years after Allport 36, 37 (Dovidio et al. eds., 2005).

30 Allport, supra, at 9.

31 Dovidio, Nature of Contemporary Prejudice, supra, at 831.

32 See, e.g., John B. McConahay, Modern Racism, Ambivalence, and the Modern Racism Scale, collected in Prejudice, Discrimination and Racism, supra, 91, 92-93.

33 Modern racism reflects the "psychological syllogism" that "Racism is bad and other beliefs do not constitute racism because these beliefs are empirical facts." McConahay, supra, at 93.

Second, John Dovidio and Samuel Gaertner coined the phrase "aversive racism" to represent "a particular type of ambivalence in which the conflict is between feelings and beliefs associated with a sincerely egalitarian value system and unacknowledged negative feelings and beliefs about blacks."[34] Though emphasizing the sincerity of the aversive racist's egalitarian commitments, Dovidio and Gaertner published empirical studies designed to test their hypothesis that the aversive racist will most likely given in to his group-based "negative feelings" when "the normative structure within the situation is weak, ambiguous, or conflicting" and that even where norms against prejudice are clear the aversive racist may "unwittingly search for ostensibly nonracial factors that could justify a negative response."[35] Dovidio and Gaertner clarified that the "negative affect" felt by aversive racists "is not hostility or hate" but may include "discomfort, uneasiness, disgust and sometimes fear, which tend to motivate avoidance rather than intentionally destructive behaviors."[36] Finally, Dovidio and Gaertner were pessimistic regarding whether aversive racists could control consciously control their negative attitudes and behaviors, concluding that "whenever aversive racists consciously monitor their behavior in interracial situations, they react in ways that consistently reinforce egalitarian self-images" and that to compel further self-examination may only lead to reverse discrimination or tokenism,[37] not to mention resentment and backlash.

Thus, already under the aversive racism model, many common sense intuitions about the relationship between prejudice and culpable behavior are disrupted. Specifically, one cannot say of aversive racists that they are purely, or even tangentially, motivated by hostility. The negative feelings they experience may be rational (e.g., one may be rational to feel guilt in response to evidence of one's arbitrary group-based social advantage; the fear that one feels of outgroup members may be reasonable in certain situations given available information). The behaviors that express their aversive racism, though real and sometimes harmful, are often not intentional in the ordinary sense that one has not acted with an intention to engage in such behaviors and may not even be aware of having done so. Perhaps most important, the aversive racism model exposes the agent as an incompetent witness to the truth of his own motivations, unable to escape his own deceptions and best-case explanations of his behavior as he struggles to maintain an egalitarian self-image.[38]

---

34 Samuel L. Gaertner & John F. Dovidio, The Aversive Form of Racism, collected in Prejudice, Discrimination, and Racism, 61, 62 (John F. Dovidio & Samuel L. Gaertner eds., 1986).

35 Gaertner & Dovidio, Aversive Form of Racism, supra at 66.

36 Gaertner & Dovidio, Aversive Form of Racism, supra, at 63.

37 Gaernter & Dovidio, Aversive Form of Racism, supra, at 85.

38 It is worth noting that Dovidio and Gaertner repeated and expanded their research in 1998-99, concluding ultimately that, while over racism may have decreased over the prior decade, aversive racism remained "more persistent." Dovidio, Nature of Contemporary Prejudice, supra, at 837.

During the same period, psychologists also began explore in more detail Allport's insight that group-based stereotypes are a normal, even inevitable, byproduct of human cognition. Following decades of research culminating in disagreement regarding whether stereotypes are the result of especially inaccurate generalizations or normal generalizations, and whether stereotypes arise from faults in the subject's reasoning or experience, the social cognition approach to stereotyping took hold, causing researchers to concentrate on the process of stereotyping rather than the content of stereotypes.[39] One prominent approach, discussed more than a decade ago by legal scholars,[40] hypothesized stereotypes to result from automatic cognitive strategies for information processing whereby the mind constructs "schemas," or knowledge structures,[41] consisting of many elements that help the mind conserve cognitive resources by processing information in a passive way. Once a category is defined, incoming information is related to that category and interpreted in accordance with expectations associated with that category.[42] Once social categories such as race-based and sex-based groups are assigned, schema theory anticipates social perceivers will exaggerate differences between groups and minimize differences within groups.[43] Furthermore, the behavior of group members is perceived in stereotyped terms, based on the perceivers associations with the group.[44] Thus, the schema theory of stereotyping disrupts the common intuition that discrimination occurs at a critical moment of judgment when the agent chooses to engage in a particular behavior or to take a particular action. Rather, the schema theory suggests that instances of perception and passive information processing may be equally important in assessing whether resulting behavior is discriminatory.

Focus on the cognitive process of stereotyping has also led to other theories that may have substantial importance for antidiscrimination law. For example, stereotyping may not always occur whenever a person is confronted by an outgroup member, and it may occur even in the absence of direct confrontation by an outgroup member, whenever the agent is primed by social stimuli. These stimuli, or primes, may be as simple as word, phrase or picture that holds some gender-specific or race-specific association. Once the prime activates the cognitive stereotyping process, it has the potential to call forth other associations the agent is disposed to make with the gender or race category.[45] This stereotyping process has various effects on information processing,

---

[39] See generally, David J. Schneider, The Psychology of Stereotyping 11-13 (2004).

[40] Krieger, Content of Our Categories, supra, at 1199-1204.

[41] Schneider, supra, at 122. See generally William F. Brewer & Glenn V. Nakamura, The Nature and Function of Schemas, in 1 Handbook of Social Cognition, 119 (Robert S. Wyer & Thomas K. Srull eds., 1984); Shelley E. Taylor & Jennifer Crocker, Schematic Bases of Social Information Processing, in Social Cognition, supra, 89.

[42] Taylor & Crocker, supra, at 97.

[43] Shelley E. Taylor et al., Categorical and Contextual Bases of Person Memory and Stereotyping, 36 J. Personality & Soc. Psych. 778, 790-91 (1978)

[44] Id. at 791.

[45] Schneider; Fiske; Bargh.

including causing the agent to fail to perceive or remember individuating information about the target and causing the agent to overemphasize negative information about the outgroup target and underemphasize similar information about her ingroup competitors.[46]

Though process driven in their approach, second wave stereotyping scholars did not neglect to address the content of stereotypes. They proposed two significant theories by having done so. First, stereotypes may be prescriptive or descriptive. In the context of sex stereotyping, prescriptive stereotypes consist of "beliefs about the characteristics that women should possess" or roles believed to be appropriate for their sex.[47] "Whereas the descriptive component of the female stereotype might consist of beliefs that women are nurturing and soft-spoken, the prescriptive component might consist of beliefs that women should be nurturing and soft-spoken."[48] Diana Burgess and Eugene Borgida hypothesize that prescriptive stereotypes may lead to workplace harassment of women who enter traditionally male-dominated occupations or negative evaluations of women seeking employment or promotion who are perceived to lack "femininity."[49] Descriptive stereotyping may be reflected in disparate impact challenges, where qualifications may be unnecessarily based on expectations regarding typical male performance.[50] However, it may also lead to discriminatory evaluations, as employers conform their perceptions of performance to their sex-based expectations of ability.

Second, the agent's associations with the category need not be "negative" in order for stereotyping to impact the accuracy of the agent's social perceptions or to provoke discriminatory behavior. Therefore, like aversive racism, sexism may reflect ambivalent attitudes and associations regarding the target group. As Susan Fiske and Peter Glick have observed in connection with sex stereotypes, automatic stereotypes may be either hostile or benevolent in nature, but paternalistic stereotyping with benevolent motivations is no less pernicious than hostile stereotyping since both rob the target of the benefits of being considered as an individual. [51] Furthermore, attitudes that begin as paternalistic, fueled by descriptive stereotypes concerning the target's need for special

---

[46] Fiske & Neuberg. Bodenhausen. Taylor et al.

[47] Diana Burgess & Eugene Borgida, *Who Women Are, Who Women Should Be: Descriptive and Prescriptive Gender Stereotyping in Sex Discrimination*, 5 Psych., Pub. Pol., & L. 665, 666 (1999); *see also* Alice H. Eagly & Steven J. Karau, *Role Congruity Theory of Prejudice Toward Female Leaders*, 2002 Psych. Rev. 573, 573-74 (2002) (describing hybrid "role congruity theory of prejudice" whereby women may face prejudice due to the perceived transgression of actual or ideal gendered behavioral expectations).

[48] Burgess & Borgida, supra, at 666.

[49] Burgess & Borgida, supra, at 667.

[50] Burgess & Borgisa, supra, at 666.

[51] Peter Glick & Susan T. Fiske, The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism, 70 J. Personality & Soc. Psych. 491 (1996).

assistance or consideration, may quickly turn hostile if the target frustrates the agent's stereotypic expectations.

The third wave of prejudice studies finds us in the present, where the familiar social objective of explaining prejudice, its impact on behavior and its persistence irrespective of the advance of egalitarian and pro-diversity ideals has merged with the cognitive science of implicit social perception. Many of the difficulties raised by second wave studies regarding common intuitive assumptions about prejudice and discrimination have been made more acute by the introduction of implicit measures that purport to assess with new accuracy the influence of bias at an unconscious level. That is, the implicit cognition model carries forward an understanding of prejudice, or bias, structured in ambivalence that was a defining feature of second wave research. Furthermore, it underscores with enhanced authority the cognitive rift between a person's conscious commitments and unconscious motivations. In anticipation that many who submit to an online version of the most popular implicit measure of racism, the IAT, will react with anxiety and disbelief upon learning that they register bias in favor of whites and against African-Americans, Harvard researchers have posted the following response to the question ("If my Black-White attitude IAT shows automatic White preference, does that mean that I'm prejudiced?"):

> Answer: This is a very important question. Social psychologists use the word "prejudiced" to describe people who endorse or approve of negative attitudes and discriminatory behavior toward various out-groups. Many people who show automatic White preference on the Black-White attitude IAT are not prejudiced by this definition. It is possible to show biases on the IAT that are not consciously endorsed, or are even contradictory to intentional attitudes and beliefs. People who hold egalitarian conscious attitudes in the face of automatic White preferences may able to function in non-prejudiced fashion partly by making active efforts to prevent their automatic White preference from producing discriminatory behavior. However, when they relax these active efforts, these non-prejudiced people may be likely to show discrimination in thought or behavior. The question of relation between implicit and explicit attitudes is of great interest to social psychologists, several of whom are doing research on that question for race-related attitudes. For more information see Banaji, Nosek & Greenwald, 2004.[52]

Thus, in a very real sense, to require that antidiscrimination law construct a definition of *discrimination* in which the latter is motivated by implicit bias is already to require that

[52] https://implicit.harvard.edu/implicit/demo/background/faqs.html#faq20 (visited on March 23, 2009). The article to which this answer refers is Mahzarin R. Banaji et al., *No Place for Nostalgia in Science: A Response to Arkes and Tetlock*, 15 Psych. Inquiry 279 (2004).

antidiscrimination law exclude prejudice from a constituent role in defining *discrimination*.

Of course, the online answer equivocates in several respects. First, it does not clarify whether the researchers themselves endorse any general definition of *prejudice* that would limit the latter to "negative" attitudes "endorsed" by the agent. In fact, at the close of the answer, the test-taker is directed to an article in which Mahzarin Banaji, Brian Nosek and Anthony Greenwald, in which they expressly distinguish between implicit and explicit prejudices, reject the notion that antipathy is a requirement of prejudice in either form, and permit endorsement to define only *explicit prejudice*.[53] Second, though the answer suggests that automatic bias may cause "non-prejudiced people . . . to show discrimination in thought and behavior," one is left uncertain as to whether non-prejudiced people are culpable for their failures to use "active efforts" to curb the effects of bias. Third, the answer introduces some questionable concepts, such as "discrimination *in thought*" and "*intentional* attitudes and beliefs."[54] One hardly knows what these are, if they even exist.

Nevertheless, whether implicit cognition researchers wish to retain some distinction for the term *prejudice* or in absorb it fully into the term *implicit bias*, two points are clear. *First*, implicit bias expands the scope of motivations that may be construed to influence discriminatory behavior, from conscious to unconscious motivations and from antipathy to ambivalence. In the role of motivating discrimination, implicit bias assumes the function historically reserved for prejudice, available to explain why behavior otherwise excusable as innocent or reasonable is culpable because of its motivation. We are encouraged to distinguish between biased, even implicitly biased, conduct and conduct that is justified. Ironically, the unconscious mechanisms of attitudinal associations and stereotyping play more of a role in suggesting a normative reason to label conduct motivated by implicit bias *discrimination* than the content of the bias itself. The apparent rational incontinence of acting on such bias itself carries a stigma that is not unlike an attribution of prejudice.[55]

Second, implicit measures hold a special significance for antidiscrimination law, and for disparate treatment jurisprudence in particular, because

---

53 See Banaji et al., No Place for Nostalgia, supra at 279-83.

54 Presumably, they mean by the latter "*conscious* attitudes and beliefs." Even if one did desire to have an attitude and devote much thinking to that desire, the subsequent coincidence of that desire with actually having the attitude would still be difficult to confirm as intentional (except perhaps in extreme cases not intended by their answer where one adapted one's attitudes through steady practice by denying satisfaction of conflicting desires and attitudes that were formerly in place so that new, adaptive attitudes emerged). As Donald Davidson has put it, "[t]he point isn't that desires and beliefs aren't ever in an agent's control, but rather that coming to have them isn't something an agent does." Donald Davidson, *Freedom to Act*, at 73, collected in ACTIONS AND EVENTS (1980).

55 Put more sharply, one is not permitted to point to the automaticity of implicit bias and claim accident rather than discrimination.

the attitudes and stereotypic beliefs they purport to measure are often at odds with the attitudes and beliefs elicited by the explicit, or testimonial, measures on which the law typically relies. Measures like the IAT crystallize this dilemma.

B. *The emergence of behavioral realism*

An impressive and growing number of legal scholars have turned to the cognitive science of prejudice in order to preserve antidiscrimination law against the judicial activism of "lay psychologists behind the bench"[56] and to purge legal doctrine of "naïve" understandings of discriminatory behavior.[57] These scholars have relied on the psychology literature to support numerous proposals.[58] Here, I am particularly concerned with a proposal that has found new life with the aid of cognitive science: that we cease to define workplace disparate treatment as *intentional* discrimination, instead tying disparate treatment to proof of "discriminatory motivation."[59] This proposal has been freshly articulated by Linda Hamilton Krieger, seconded by several others, and is widely recognized to be an intellectual descendant of seminal work done by Charles Lawrence III on "unconscious racism"[60] and David Oppenheimer on "negligent discrimination."[61]

Legal scholars have long criticized the requirement that a plaintiff may sustain a claim of unlawful discrimination only on a showing that the defendant *intentionally* discriminated. Two species of this criticism predominate. The first is the position that discrimination based on a requirement of intentionality is normatively inadequate; rather, how we define discrimination should primarily revolve around what we take to be the harm associated with discrimination – that is, status-based inequality.[62]

---

[56] Krieger, The Intuitive Psychologist Behind the Bench: Models of Gender Bias in Social Psychology and Employment Discrimination Law.

[57] Greenwald & Krieger, Implicit Bias: Scientific Foundations, supra, at 946. Also Banaji & Kang, Fair Measures ("hypocrisy and self-deception").

[58] Use of IAT to screen for employment or government service. (Ayres) Use of increased affirmative action programs (Banaji & Kang). Creation of regulatory regime

[59] Krieger & Fiske, Behavioral Realism, at __.

[60] Charles R. Lawrence III, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 Stan. L. Rev. 317 (1987).

[61] David B. Oppenheimer, *Negligent Discrimination*, 141 U. Pa. L. Rev. 899 (1993)

[62] This criticism is particularly well-suited as a challenge to the constitutional doctrine of equal protection, where evidence of a status-based discriminatory impact is not constitutionally salient unless it is also circumstantial evidence of a discriminatory purpose. *See, e.g.*, Owen M. Fiss, *Groups and the Equal Protection Clause*, 5 Phil. & Pub. Aff. 107 (1976) (arguing that the Equal Protection Clause implies a group-disadvantaging principle, prohibiting the perpetual subordination of disadvantaged groups; David A. Strauss, *Discriminatory Intent and the Taming of* Brown, 56 U. Chi. L. Rev. 935, 940-46 (1989) (discussing equality-based alternatives discarded by the Supreme Court when it formalized the discriminatory purpose requirement under equal

The second criticism is the position that the intentionality requirement misstates the nature of discriminatory behavior, consequently raising undue evidentiary burdens and impairing the remedial objectives of antidiscrimination law. This empirical argument questions whether the intentionality requirement permits antidiscrimination law to address subtle, or unconscious, discrimination.[63] Each of these positions takes a foothold in the fact that the intentionality requirement reflects a combination of normative commitments and empirical assumptions regarding the nature of discrimination, its causes and its consequences. Behavioral realism represents a strong turn toward the empirical argument by aligning legal interpretation with the science of implicit social cognition.

Over a decade ago, Linda Hamilton Krieger introduced legal scholars to the cognitive psychology literature on prejudice and stereotyping, in an effort to explain why disparate treatment jurisprudence "while sufficient to address deliberate discrimination prevalent in an earlier age, is inadequate to address the subtle, often

---

protection doctrine). It is less relevant to the evaluation of statutory regimes that contain an alternative "disparate impact" theory of discrimination, permiting plaintiffs to establish liability on the basis of employment practices causing disparate impact on the basis of social status even in the absence of intentional discrimination. *See Griggs v. Duke Power Co*., 401 U.S. 424 (1971) (establishing availability of disparate impact discrimination claim under Title VII); *see also* 42 U.S.C. § 2000e-2(k) (codifying disparate impact discrimination claim and overruling modification of the claim stated by the Supreme Court in *Wards Cove Packing Co. v. Antonio*, 490 U.S. 642 (1989)). Therefore, the more discerning question is not *whether* the plaintiff should be required to establish intentionality (i.e., she will not always be required to do so) but *when*.

63 *See, e.g*., Charles R. Lawrence III, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 Stan. L. Rev. 317 (1987) (using psychoanalytic theory to argue that the unconscious nature of prejudice renders the doctrinal intentionality requirement descriptively inaccurate and incompetent to isolate a broad range of discriminatory behaviors); *see also* Martha Chamallas, *Deepening the Legal Understanding of Bias: On Devaluation and Biased Prototypes*, 74 S. Cal L. Rev. 747, 753 ("[A]ntidiscrimination law is inadequate because it targets mainly intentional discrimination, missing the more prevalent contemporary forms of bias that are often nondeliberate or unconscious."); David B. Oppenheimer, *Negligent Discrimination*, 141 U. Pa. L. Rev. 899 (1993) (using tort theory and social psychology to argue that an additional category of "negligent discrimination" is necessary to permit Title VII to target "unconscious discrimination"); David A. Strauss, *The Law and Economics of Racial Discrimination in Employment: The Case for Numerical Standards*, 79 Geo. L.J. 1619 (1991) (arguing that, because overt racial discrimination based on animus is largely absent in the modern workplace, the remedial objectives of Title VII can best be achieved by requiring employers to meet explicit numerical standards); Susan Sturm, *Second Generation Employment Discrimination: A Structural Approach*, 101 Colum. L. Rev. 458, 460 (2001) ("Cognitive bias, structures of decisionmaking, and patterns of interaction have replaced deliberate racism and sexism as the frontier of much continued inequality.").

unconscious forms of bias that Title VII was also intended to remedy."[64] Krieger's watershed contribution to the legal literature on discrimination set the stage for the current cognitive turn in how we theorize about antidiscrimination law. Among other important contributions, Krieger's early article describes how our understanding of disparate treatment as intentional discrimination requires the factfinder to "find that the decisionmaker intended to discriminate or that no discrimination occurred."[65] Recounting substantial evidence from the psychology literature that "normal" unconscious cognitive processes "tend to bias intergroup perception and judgment," Krieger concludes that "there now exists a fundamental 'lack of fit' between the jurisprudential construction of discrimination and the actual phenomenon it purports to represent."[66]

Identification of a fundamental mismatch between law and the empirics of the phenomena that it seeks to regulate is now a familiar refrain of the behavioral realist movement. The introduction of implicit bias theory into legal analysis and policy is therefore viewed as a necessary corrective measure. Christine Jolls and Cass Sunstein have described behavioral realism as an approach whereby, "in formulating and interpreting legal rules, legislatures and courts should pay close attention to the best available evidence about people's actual behavior" and they add that this approach "should not be controversial."[67] Linda Hamilton Krieger and psychologist Susan Fiske have taken the definition one step further, observing that judges instantiate behavioral theories as a part of legal doctrine whether or not they acknowledge doing so. Krieger and Fiske target their criticisms against so-called "common sense" or "intuitive" psychological notions that they see manifested in disparate treatment doctrine and identify as having no empirical basis, boasting that a "psychologically trained eye can spot these intuitive psychological theories all across Title VII's doctrinal landscape."[68]

Regarding disparate treatment jurisprudence, behavioral realists argue that the intentionality requirement mischaracterizes the nature of discrimination in modern society, as this is often the result of implicit bias operating beyond the actor's conscious awareness.[69] They further argue that the concept of intentional discrimination cannot

---

[64] Linda Hamilton Krieger, The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity, 47 Stan. L. Rev. 1161, 1164 (1995).

[65] Id. at 1170.

[66] Id. at 1217.

[67] Christine Jolls & Cass R. Sunstein, *The Law of Implicit Bias*, 94 Cal. L. Rev. 969, 972 (2006).

[68] Linda Hamilton Krieger & Susan T. Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 Cal. L. Rev. 997, 1006 (2006).

[69] *See, e.g.*, Linda Hamilton Krieger, The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity, 47 Stan. L. Rev. 1161 (1995); Linda Hamilton Krieger, The Intuitive Psychologist Behind the Bench: Models of Gender Bias in Social Psychology and Employment Discrimination Law, 60 J. Soc. Issues 835 (2004); Linda Hamilton Krieger, Naturalizing Disparate Treatment Theory,

account for the effect that implicit bias may have in provoking discriminatory behaviors because it necessarily implies that the employer must be aware of its biased reasons for acting and must act with the intention to fulfill those reasons.[70] As Krieger and Fiske have stated, "one can be 'nonprejudiced' as a matter of conscious belief and yet remain vulnerable to the subtle cognitive and behavioral effects of implicit stereotypes."[71] Thus, behavioral realism purports to show a mismatch between, on the one hand, the empirical assumptions of disparate treatment doctrine manifested as "intuitive psychological theories"[72] and, on the other hand, the empirical realities of the discriminatory practices that the doctrine aims to address.[73] And it purports to do that by using implicit bias as a means to show that those who outwardly appear "nonprejudiced" or who do not even know themselves to be prejudiced may otherwise suffer from implicit bias – that is, their actions may be motivated by implicit prejudice or implicit stereotypes.

Consider the paradigm case of an employer who believes that his decision to terminate the plaintiff is based on material if only relative differences between her performance and the performance of other peer employees who were not terminated. The employer may testify sincerely and credibly that the decision was based on legitimate, nondiscriminatory reasons related to the differences in job performance. However, the reason that certain differences became salient to the employer may not have anything to do with the objective weight assigned to different aspects job performance. Rather, the employer may have been motivated to view those differences as salient because of the plaintiff's protected class status. The employer is therefore wrong about his reasons and, according to implicit cognition theorists, in a very ordinary sense. Put more succinctly, the employer is mistaken in believing that his articulated reasons for acting are in fact what motivated his action. For the reasons discussed in Part I.A., implicit bias theory

---

collected in NYU Selected Essays on Labor and Employment Law, Vol. 3 (Mitu Gulati & Michael Yelnosky eds., 2007); Krieger & Fiske, Behavioral Realism in Employment Discrimination Law, supra note __. *See also* Gary Blasi, Default Discrimination: Law, Science, and Unintended Discrimination in the New Workplace, collected in NYU Selected Essays on Labor and Employment Law, Vol. 3 (Mitu Gulati & Michael Yelnosky eds., 2007); Jolls & Sunstein, The Law of Implicit Bias, supra note __.

70 See, e.g., Krieger & Fiske, Behavioral Realism in Employment Discrimination Law, supra note __, at __.

71 See, e.g., Krieger & Fiske, Behavioral Realism in Employment Discrimination Law, supra note __, at 1033.

72 See, e.g., Krieger & Fiske, Behavioral Realism in Employment Discrimination Law, supra note __, at 1006.

73 Given the nature of this critique, one of the more curious points about the version of behavioral realism practiced by Linda Hamilton Krieger is that it represents that it is not participating in an argument about norms but only in an argument about empirical descriptions of social practices. Thus, she characterizes the intentionality requirement as an empirical claim about the causes of discrimination – the result of judges acting as "intuitive psychologists" – rather than as a normative claim about what makes certain forms of discrimination wrongful or not legally excusable. See, e.g., Krieger, Intuitive Psychologist, supra note __, at __; Krieger & Fiske, supra note __, at __.

promises to explain this case as a case of discrimination by identifying implicit bias as the employer's true motivation. However, implicit bias theory cannot tell us why we should prohibit the full range of discriminatory behaviors susceptible to its diagnosis or why we cannot decide upon a definition of *discrimination* that transcends the cognitive account of what decisions are biased and thus unjustified.

The behavioral realists owe much to Charles Lawrence's pioneering article, *The Id, the Ego, and Equal Protection*, in which he uses psychoanalytic theory to describe why the intentionality requirement of equal protection doctrine is descriptively inaccurate and instrumentally incapable of identifying a significant range of discriminatory behaviors.[74] Many behavioral realists have acknowledged that debt and style their own work as a kind continuation – even a scientific upgrading – of Lawrence's thesis based on implicit bias research.[75] However, Lawrence's argument was based in part on the realization that "the injury of racial inequality exists irrespective of the decisionmaker's motives,"[76] and this is a very different position than the behavioral realists who, following the implicit cognition theory, identify implicit motivations as means to make causal claims about discriminatory behaviors. Having joined itself to implicit cognition theory, it is no surprise that behavioral realism takes unconscious motivation to be its touchstone. Indeed, Krieger and Fiske ultimately propose to amend disparate treatment doctrine by replacing the requirement of intentional discrimination with a requirement of "discriminatory motivation" which may be demonstrated by "thoughtless" or "unwitting application" of social stereotypes.[77]

The behavioral realists reject the intentionality requirement for two distinct types of practical purposes.[78] *First*, Krieger and Fiske aim specifically to

---

74 Lawrence, supra note __, at __.

75 [CITE]

76 Lawrence, *supra* note __, at 320. Of course, Lawrence had a different complaint to lodge against the constitutional doctrine, due to the absence of any disparate impact claim under equal protection law, than the behavior realists are permitted to have against Title VII. Given that Title VII makes a disparate impact claim available to litigants, we might wonder why Krieger and Fiske do not attack the insufficiency of the doctrine that does not contain an intentionality requirement rather than attacking disparate treatment doctrine because it does have that element.

77 Krieger & Fiske, Behavioral Realism in Employment Discrimination Law, supra note __, at 1058-61. Krieger and Fiske state that "[w]hat changes under a behavioral realist interpretation of [Title VII] is the set of inferences that can reasonably be drawn from these species of evidence and what exactly 'discriminatory motivation' means, as an essential element of the plaintiff's disparate treatment claim." Id. at 1060. In addition, Krieger has separately purported to join other scholars in the "law and social cognition school" who "generally advocate a causation-based, rather than an intent-based understanding of disparate treatment discrimination," Krieger, Naturalizing Disparate Treatment Theory, supra note __, at 509.

78 Here, I mean purposes that involve a direct impact to the outcome of cases or the manner in which they are litigated.

eradicate lower court doctrines that have developed to structure particular factual inquiries. Under the pretext analysis established by the Supreme Court in *McDonnell Douglas v. Green*,[79] the plaintiff must first establish a prima facie case, supporting a presumption of intentional discrimination. The defendant may rebut the presumption by proffering evidence of a legitimate, nondiscriminatory reason, and, if he does so, the plaintiff must prove that the proffered reason was a pretext and "that discrimination was the real reason."[80] Lower courts have sometimes sought to streamline this burden-shifting process by positing additional judicial inferences. For example, the Fourth Circuit follows a "same actor inference" rule stating a strong inference against a finding of discrimination in termination cases "where the hirer and the firer are the same individual and the termination of employment occurs within a relatively short time span following the hiring."[81] I agree with Krieger and Fiske that the same actor inference should be rejected under Title VII.[82] But the same actor inference is not necessitated by the intentionality requirement, and discarding intentionality is a disproportionate response to occasional lower court departure from Supreme Court precedent.[83] In short, having the purpose of curtailing this sort of doctrinal drift is no reason to reject the architecture established by Supreme Court precedent.

Second, Krieger and Fiske are joined by other scholars in positing a new paradigm case of discrimination in which the motivations of the defendant cannot be known except based on evidence of their implicit bias. For example, Jolls and Sunstein posit a case in which the employer must decide whether to promote a white employee, Jones, or a black employee, Smith. The employer "thinks that both employees are excellent, but it chooses Jones on the bases of a 'gut feeling' that Jones would be better for the job" and he "thinks that 'Jones is a better fit.'" Otherwise, the employer did not consciously think of race in making its decision, but "Smith would have been chosen if

---

79 411 U.S. 792 (1973).

80 *Id*. at 510-11 & n.4 (internal quotation marks and citation omitted). For a more developed account of disparate treatment doctrine, see *infra* Part II.B & C.

81 *Proud v. Stone*, 945 F.2d 796, 797-98 (4th Cir. 1991) (developing rule under the ADEA). Not all circuits observe this rule, and some that do dilute it significantly. [CITE cases] *See also* Krieger & Fiske, Behavioral Realism in Employment Discrimination Law, supra note __, at 1047-48 (discussing same actor inference as applied by several circuits).

82 Not only, as Krieger and Fiske rightly argue, are employers affected by situational cues arising from the specific contexts of a decision to hire or terminate an employee, but lower courts also should not be cavalier about adding judicial inferences at the summary judgment or at trial that affect the substantive rights of the parties and are not otherwise authorized by Supreme Court doctrine.

83 Krieger and Fiske also reject another lower court doctrine, the "honest belief rule," for similar reasons. They are wrong about the origin and essential meaning of the honest belief rule, and so I reject their criticisms of it. *See infra* note [37]. But if their interpretation of the doctrine would true, it would represent no more of a reason to reject the intentionality requirement established by the Supreme Court than the same actor inference.

both candidates had been white."[84] The assumption here is that the intentionality requirement would make it difficult for Smith to prove a case of disparate treatment discrimination, because he will have difficulty rebutting the employer's contention that race did not enter into his decision-making process when he assessed Jones's superior "fit." Similarly, Krieger has posited that an employer's unequal treatment of women based on stereotypes about their roles as mothers would state a claim of disparate treatment if the plaintiff is permitted to prove discriminatory motivation based on "spontaneous, uncorrected forms of implicit bias . . . but not if motivation is equated with self-conscious intentionality."[85]

What is the essential feature that unifies these cases into a particular species of disparate treatment case? That the defendant's proffered legitimate, nondiscriminatory reason is true so far as the defendant is aware. That is, the employer proffers an honest account of its reasons from its own point of view, or the point of view of the person(s) responsible for taking an employment action adverse to the plaintiff. Furthermore, this paradigm case assumes that the plaintiff cannot demonstrate that the defendant's "honest" reason is a pretext, in that she is unable to provide persuasive evidence that the reason is a deception aimed to conceal the defendant's true discriminatory motivation. However, if the doctrine will permit the plaintiff to introduce evidence of the defendant's implicit bias, she may be able to demonstrate that the defendant's "honest" reason is not its true reason – that the defendant, or its agent(s), has mistaken his conscious reason for the cause of his action.

I agree that considering this paradigm case helps us to refocus antidiscrimination law away from cases where evidence of conscious prejudice is available.[86] In the American workplace, such cases have long been in decline.[87] But I disagree with the behavioral realists' conclusion that these cases do not show intentional discrimination. To the contrary, each shows an adverse employment action, taken in an intentional manner, for discriminatory reasons of which the defendant is not completely, or perhaps not at all, aware. Cognitive psychology may assist courts to understand how implicit bias may motivate deliberate and intentional discriminatory behaviors, even without the employer's awareness that that any bias influenced its decision. This seems to me to present an opportunity to clarify our description of intentionality under disparate treatment doctrine to explain how it is compatible with evidence of implicit bias and why this is the right interpretation both normatively and descriptively.

The more fundamental problem with seizing on any particular paradigm case is that, though it may release us from certain artificial limitations affecting how we view the doctrine, it is bound to reinforce other limitations or to impose new ones. The behavioral realists effectively place social science in something of a superordinate

---

84 Jolls & Sunstein, *The Law of Implicit Bias*, *supra* note __, at 970.

85 Krieger, *The Intuitive Psychologist Behind the Bench*, supra note __, at 840.

86 My agreement on this point persuades me to applaud several contributions of behavioral realism. *See infra* Part I.B.

87 [CITE]

relationship to antidiscrimination law's construction of what constitutes wrongful discrimination. Krieger and Fiske offer a "psychologically trained eye" as a "judicial corrective" to lay psychological theories that reflected in judicial opinions. Of course, it is no surprise that the eye can see what it is trained to see. Sometimes, that is its limitation.

In Part II, I will consider a different type of case from these paradigm cases, in order to examine what role an employer's benign or benevolent motivations ought to have in excusing disparate treatment. Implicit bias research provides certain benefits for the adjudication of cases lacking evidence of conscious motivation, but it has relatively little to contribute to our analysis of cases where the plaintiff can demonstrate that her disparate treatment was because of her social status but cannot disprove the defendant's proffered benign or benevolent reason. To be sure, one may respond that implicit bias research permits the factfinder to ignore the defendant's proffered reason on the ground that he likely does not know his own true motivations. As I demonstrate in Part II, disparate treatment doctrine focuses on the defendant's motivations only to assess whether it took a particular action because of the plaintiff's status; the defendant's motivations are not relevant to excuse disparate treatment otherwise demonstrated to have occurred because of the plaintiff's status. For example, if a supervisor takes an adverse employment action against the plaintiff such as delaying her consideration for a promotion or denying her complete and honest feedback because she is a woman, disparate treatment doctrine does not excuse the employer from liability just because the supervisor acted with a benevolent affect or a specific intention to advantage the plaintiff in some way.[88]

Finally, what if the "mismatch" targeted by the behavioral realists is defined not by a contradiction between the doctrine's empirical assumptions and psychological research, but instead by a contradiction between the doctrine's normative commitments and this research?[89] What authority does this research have to unseat or force amendment of the law's normative commitments, and what new normative outlook does it command? As with so many other critiques of disparate treatment doctrine, the proposed path to reform – here, to the correction of this mismatch – requires, that we jettison the intentionality requirement.[90] But, even if in its present form the

---

88 *See infra* Part II.B.

89 Behavioral realists have not developed a coherent account of the normative commitments that underlay the intentionality requirement, nor are they convincing in their efforts to describe intentionality as a descriptive rather than normative element of the doctrine. I will discuss in greater detail below what I believe are the important features of the normative commitments reflected in the intentionality requirement. I will say here in summary that they include commitments (i) to eradicate unequal treatment, deliberate rather than accidental, characterized by the defendant's exercise (or capacity to have exercised) control over the action, and (ii) to preserve of the employer's legitimate business judgment. *See infra* notes 99-101 and accompanying text.

90 However, to what end is not entirely clear. Linda Hamilton Krieger has proposed that the intentionality requirement be replaced by proof of a "discriminatory motivation" (which may include implicit bias), see e.g., Krieger, The Content of Our Categories, supra note

intentionality requirement were incompatible with implicit bias evidence – which it is not – the instruction to jettison the requirement is not compelling if the requirement reflects something more than just a descriptive theory of discrimination. If the requirement reflects instead a commitment to define discrimination in a particular way, is it not incumbent on legal scholars to use new empirical understandings to fulfill the law's normative commitments or else demonstrate why those commitments, in and of themselves, ought to be rejected or amended?

B. *What the behavioral realist account contributes to our understanding of the doctrine*

Though I take issue with certain aspects of behavioral realism, many of its contributions are profitably clarifying and have come not a moment too soon. Scholars and courts have long complained that disparate treatment doctrine has over time become incapable of addressing evolving forms of "subtle" discrimination.[91] In developing disparate treatment doctrine, the Supreme Court specifically intended that this not be the case.[92] However, over time, the case law muddied the concept of disparate treatment by implicitly equating intentional discrimination with proof of "discriminatory motive" and "discriminatory intent,"[93] thereby suggesting that it is the plaintiff's burden to prove the defendant's state of mind and presenting a context vulnerable to the sort of doctrinal drift that I discuss above.[94] Simultaneously, employers have become increasingly adept at concealing evidence of illicit motivations and producing evidence that challenged employment actions were based on nothing other than legitimate, nondiscriminatory business reasons. Now, many see difficulty in proving disparate treatment as endogenous to the doctrine, reflecting the unreasonableness of the intentionality requirement rather

---

__, at __; Krieger & Fiske, Behavioral Realism in Employment Discrimination Law, supra note __, at 1060 ("What changes under a behavioral realist interpretation of [Title VII] is the set of inferences that can reasonably be drawn from these species of evidence and what exactly 'discriminatory motivation' means, as an essential element of the plaintiff's disparate treatment claim."), and she has also purported to join other scholars in the "law and social cognition school" who "generally advocate a causation-based, rather than an intent-based understanding of disparate treatment discrimination," Krieger, Naturalizing Disparate Treatment Theory, supra note __, at 509. See also Blasi, supra note __, at 4-5 (arguing that the current intentionality doctrine should be supplemented by a regulatory regime that facilitates the development and implementation of "best practices" against workplace discrimination).

[91] *See, e.g.*, *Aman v. Cort Furniture Rental Corp.*, 85 F.3d 1074, 1082 (3d Cir. 1996) ("Discrimination continues to pollute the social and economic mainstream of American life, and is often simply masked in more subtle forms. It has become easier to coat various forms of discrimination with the appearance of propriety, or to ascribe some other less odious intention to what is in reality discriminatory behavior."). *See also supra* note __ and accompanying text.

[92] *See McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 801 (1973) ("Title VII tolerates no racial discrimination, subtle or otherwise.").

[93] [CITE cases]

[94] *See supra* note __ and accompanying text. [EXAMPLES]

than resulting merely from evidentiary asymmetries between employers and victims of discrimination.[95]

The logic of implicit bias has certain similarities and tensions with legal notions of circumstantial proof, particularly circumstantial proof of intent. Both make room for the possibility that a person's state of mind may be inferred from evidence that is less direct than a confession or recording of that state of mind. However, legal methods traditionally privilege directly questioning the subject and then challenging the veracity or credibility of his testimony. These methods are often inadequate in circumstances where the subject sincerely or plausibly testifies to having acted on nondiscriminatory reasons (as in the paradigm case posited by the behavioral realists),[96] for in such circumstances one may not be able to rule out the existence of other illicit motivating factors. Implicit bias research may help plaintiffs to overcome some of the hurdles associated with the disparate treatment burden of proof, and it may convince courts to resist imposing new judicial inferences that curtail plaintiffs' access to such strategies for proving their claims.

I therefore agree that implicit bias research advances our legal understanding of discrimination in three significant ways. *First*, this research provides us with a model to test the efficacy of legal doctrine in its efforts to identify and remedy discrimination. This is the general claim that Professors Krieger and Fiske make about behavioral realism – that it forces judges and lawmakers to justify legal doctrine according to empirical understandings regarding their juridical and regulatory subjects. But I would amend their claim somewhat. Krieger and Fiske boldly reject that notion that normative aims and not empiricism should guide the development of substantive legal doctrine.[97] However, there is a very real sense in which the law must define wrong before it may undertake to discern or remedy wrong, and Krieger and Fiske do not deny this.[98] Therefore, it seems that one particularly good use of the clear empirical picture provided by implicit bias research is that it provides a framework against which we may assess whether a particular court has deviated from the normative commitments of disparate treatment doctrine by imposing its own empirical presuppositions.[99]

95

96 See supra note __ and accompanying text.

97 Krieger & Fiske, *Behavioral Realism in Employment Discrimination Law*, *supra* note 69, at 1005-06.

98 *See infra* note __ and accompanying text.

99 I take this to be exactly what Krieger and Fiske have done in showing that the circuit court doctrines of the "same actor inference" and "honest belief rule" are deviations from established statutory and Supreme Court doctrines based on faulty premises about cognition and prejudicial predisposition. Krieger & Fiske, Behavioral Realism in Employment Discrimination Law, supra note __, at __ - __. But I view the critique of the honest belief rule as the less successful of the two because there is an important normative distinction between excusing an employer for taking adverse action against a plaintiff based on the honest belief that the plaintiff was inferior in some way to other

Second, implicit bias research expands our understanding of what may be probative circumstantial evidence that a defendant acted "because of" a plaintiff's protected class characteristic. For example, if we want to consider whether an employer unlawfully discriminated in terminating a female employee, we are encouraged to ask not only whether the proffered reason (*e.g.*, poor job evaluation scores) is true, but also whether that reason was deployed differently than it would have been for a male employee (*e.g.*, for whom leadership, profitability or some other factor may have been given greater weight) and whether the information relied upon was developed or acquired in a biased manner (*e.g.*, that women and men with objectively comparable performance may have been evaluated differently). Thus, implicit bias research counsels us to ask not only what the reason for an action was at the moment that the action was taken (*e.g.*, perhaps the plaintiff was terminated because she had the lowest job evaluation scores) but also whether the choice to rely on particular criteria was biased or whether the development of information related to that criteria was biased.[100]

Third, implicit bias research significantly expands traditional understandings of prejudice – which typically equated prejudice with conscious animus and defined it as a pathology rather than as a common or even basic condition of the way that human beings process information. Gordon Allport's influential definition of prejudice as "thinking ill of others without sufficient warrant,"[101] at once affirms and rejects the traditional understanding, for Allport also stressed the "normality of prejudgment."[102] Among contemporary psychologists, the normality, or automaticity, of prejudice and stereotyping has taken on increased importance. For example, Professor Fiske has remarked that "automatic categorization and automatic associations to categories are the main culprits in the endurance of bias."[103] Irene Blair has written, with respect to actions motivated by implicit biases, "[p]eople may often not be aware of what they are doing, they might even intend to be doing something else."[104] Mahzarin Banaji,

---

employees and excusing an employer for taking adverse action when the employer mistakenly believes that the plaintiff acted criminally or in violation of company policy (unless the employer's mistaken belief was caused in some way be the plaintiff's protected class characteristic). It seems to me that there is nothing normatively disagreeable about a rule that a mistaken determination of fact does not in itself demonstrate intentional discrimination.

[100] However, just how far the law will extend in adopting the modes of motivational testing or the dispositional presumptions that psychological researchers have tied to particular situational cues (e.g., concerning questions like whether a person's work environment will predispose them to have or to act on certain biases) is a normative question that implicit bias research cannot answer.

[101] Gordon Allport, The Nature of Prejudice 7 (1954).

[102] Id. at __.

[103] Susan T. Fiske, *Stereotyping, Prejudice, and Discrimination*, collected in THE HANDBOOK OF SOCIAL PSYCHOLOGY 357, 386 (Daniel Gilbert et al. eds., 1998).

[104] Irene V. Blair, *The Malleability of Automatic Stereotypes and Prejudice*, 6 Per. Soc. Psych. Rev. 242 (2002).

Brian Nosek and Anthony Greenwald have concluded that explicit prejudice is distinct from implicit prejudice in that a person may have implicit prejudices without demonstrating any explicit prejudice; rather, what matter is whether people "are able to function in nonprejudiced fashion" through "active efforts" to prevent implicit bias from "producing discriminatory behavior."[105] The notion that implicit bias operates beyond the agent's conscious awareness has thus expanded the understanding of prejudice commonly discussed by social and cognitive psychologists.[106] In addition, contemporary research rejects the notion that the stereotypical beliefs that underlie prejudicial, or biased, attitudes must themselves hostile. For example, Professors Fiske and Peter Glick have observed that automatic stereotypes that may be linked to discriminatory behavior may be either hostile or benevolent in nature, since the harm of stereotyping is that the target is either adjudged to conform to a set of preconceived notions (despite their individual attributes) or else is punished by negative attitudes as a consequence of their failure to conform.[107] However, behavioral realism may miss a greater potential to advance the remedial aims of antidiscrimination law to the extent that it expands doctrinally salient notions of illicit motivation from only explicit to implicit prejudice, simply adding one species within the same genus of *negative* attitudes and beliefs. The concept of implicit bias does not lose this value if it we allow it to describe phenomena that transcend prejudice to include other forms of mental causation that have not been infused with the same value judgment, though they may result in similar discriminatory behavior.

Finally, that we may find such substantial benefits in behavioral realism's uses of implicit bias research does not resolve for us the question whether its rejection of the intentionality requirement may undermine disparate treatment doctrine by threatening to confer liability on a basis that fundamentally lacks coherence or abandons core principles of rationality. Behavioral realism depicts itself as a "jurisprudential corrective"[108] to an "intuitive" approach to defining discrimination, which I call the *rationalist* approach. The rationalist approach is a kind of foil for behavioral realism,[109] and features of it appear across many articles in the literature. It comprises a set of assumptions about intentional discrimination (and, indeed, any intentional conduct) that are fundamentally incompatible with implicit bias research. Specifically, rationalists assume that, for discrimination to be intentional, the actor must be consciously aware of his prejudicial motivations as reasons for his action. In other words, he must intend to discriminate, and this intention must be based on his "taste for discrimination"[110] or

---

105 Banaji, Nosek & Greenwald, *No Place for Nostalgia*, *supra* note __, at 281.

106 See, e.g., Banaji & Greenwald, Implicit Stereotyping and Prejudice __, collected in The Psychology of Prejudice (Zanna & Olson eds.).

107 Peter Glick & Susan Fiske, Ambivalent Sexism.

108 Krieger & Fiske, *supra* note **Error! Bookmark not defined.**, at [1006].

109 Not to be confused with the "anti-scientific" foil that is also prominent. See, e.g., Kang & Banaji

110 GARY BECKER, THE ECONOMICS OF DISCRIMINATION 39-40 (2d ed., 1971) (discussing intensity of employer preferences to engage in discrimination rather avoiding the

prejudice. Furthermore, this taste for discrimination may be understood as a disposition (i.e., a relatively consistent "intensity" of attitude)[111] that we may expect to extend across social transactions.[112] Rationalists deny that an actor's prejudicial intentions or motivations can be evaluated simply as a set of probabilistic inputs or cognitive flirtations with prejudice; rather, the actor must actually endorse such attitudes in order to be judged as having acted on them.[113]

Admittedly, implicit bias research is controversial because it upends common intuitions about discrimination as a form of aberrant behavior motivated by prejudice. Instead, implicit bias research demonstrates that individuals often unconsciously harbor negative attitudes toward members of particular groups, that such attitudes are relatively common and that they may influence behavior without the individual's having become conscious of them. This sharply contradicts the *rationalist* view,[114] which holds that intentional discrimination requires the formation of an intention to act on a prejudicial belief or attitude with the objective to harm another because of , and not in spite of, her membership in a particular social group. Following what action theorists have termed the "simple view"[115] of intentional action, we might call the

---

forfeited profits and other marginal costs that may arise from discrimination); see also Krieger & Fiske, supra note **Error! Bookmark not defined.**, at __ (discussing Becker).

111 Krieger & Fiske, supra note **Error! Bookmark not defined.**, at __ (describing Becker's view as "dispositionism").

112 Krieger & Fiske, supra note **Error! Bookmark not defined.**, at __.

113 Cite to Krieger & Fiske; Mitchell & Tetlock; Arkes & Tetlock.

114 I want to suggest this view as a foil to the behavioral realists. For example, a rationalist examining disparate treatment doctrine will seek confirmation that intentional discrimination requires the formation of an intention to act on a prejudicial belief or attitude with the objective to harm another because of her membership in a particular social group. A *realist* will look at the doctrine and complain that the intentionality requirement is its weakness, because prejudice may indeed include biases that escape the agent's conscious awareness while the intentionality requirement threatens to deny the salience of such unconscious biases when determining whether a defendant's conduct constitutes discrimination.

115 CITES [Bratman, McAdams, Mele, Nadelhoffer]. Action theorists, following Michael Bratman, call this the "simple view" of intentional action: that an agent *A* to intentionally *x*, he must have intended to *x*. And this, of course, demands that *A* be correct in his understanding of what kind of action *x* is. Thus, Fred McAdams has opined that "[t]he Simple View of the relation between intention and intentional action is the commonsense view that a person does an action [*x*] intentionally only if he intends to do [*x*]." *Intention and Intentional Action: The Simple View*, 1 MIND & LANGUAGE 281, 283 (1986). But the simple view becomes quickly complicated if we ask whether *A* intended the constituent components that make up *x*, or any foreseeable "side effects" of *x*-ing, or whether *A* intentionally *x*-ed when he acted intending to *x* (and did indeed *x*) but he believed (mistakenly) that he was *y*-ing, or whether *A* intentionally *x*-ed when he thought his reason for *x*-ing was *q* even though (against his better judgment) his real reason for *x*-ing was *p* such that if not *p* then he would not have *x*-ed.

rationalist position the *simple view of intentional discrimination*, and it has a certain popular appeal. By contrast, the behavioral realists advance a theory of *implicit discrimination*[116] permitting discrimination to be proved on a showing that it was motivated by implicit bias and without requiring that the act of discrimination be intentional. This theory is reflected by the behavioral realists' paradigm case, in which the employer's proffered legitimate, nondiscriminatory reason is honestly believed by the agent responsible for the discriminatory action, and yet his true reason is that he was motivated by implicit bias.

It is ironic, therefore, that the behavioral realists do not subject the concept of intentionality to the same sort of cognitive or analytical testing to which they subject the concept of discrimination. For example, Greenwald and Krieger assert that "an intention to act is *conscious* if the actor is aware of taking an action for a particular reason."[117] They further explain that the innovation of implicit bias theory is that, while "deceitful actor[s]" may lie about their reasons for acting, they yet know the truth of those reasons; but an agent motivated by implicit bias may lack awareness or control over the true motivations of their actions. This all seems perfectly accurate. Left unexplained is whether there is, or could be, any difference between conscious and unconscious intentions,[118] or whether an action for which an agent is aware of having a reason ceases to be intentional just because the agent was ultimately was influenced to commit the action by some unconscious motivation.[119] Thus, intentionality is always discussed in the same simple, reductive terms favored by rationalists, with no sense that *intentional* may have many meanings and may be explored in different ways.

In the Part that follows, I shall offer an explanation as to why disparate treatment doctrine is confined neither by a requirement of prejudice nor by a requirement of conscious intent.

---

116 The behavioral realists' paradigm case is an illustration of *implicit discrimination*.

117 Greenwald & Krieger, supra, at 946.

118 Generally rejected as an oxymoron by the philosophical literature.

119 A more difficult question. Certainly if the definition of the action turns on its meaning, and its meaning turns on the agent's intention, then acting for some motivation other than the proper reason strips the action of its meaning and its status or name. But not all actions are defined by the meaning we attach to them. Discrimination seems not to be. Knowing an agent's state of mind under disparate treatment doctrine only aids to identify actions as discriminatory by identifying the plaintiff's protected class status as the cause of her adverse treatment.

II. How disparate treatment doctrine defines "intentional discrimination"

A. *The discrimination puzzle*

I have complained that the paradigm case of *implicit discrimination* proposed by the behavioral realists, while heuristically productive, is nevertheless incomplete. I propose to view the doctrine through a different type of hypothetical, one that allows us to put into perspective the intermittent nature of the salience of the defendant's motivations.

Consider the case of African-American attorney Emma Vaughn.[120] Vaughn was employed as an in-house contract analyst by a major U.S. corporation. On returning from maternity leave, Vaughn was informed by her immediate supervisor, *X*, that her productivity was low and that she had been entertaining too many visits from African-American coworkers in her office. Vaughn reported the conversation to her department manager, *Y*, who stated that he was aware of *X*'s criticisms, repeated those criticisms, and then warned Vaughn that she "was allowing herself to become a black matriarch" within the corporation with negative consequences for her own productivity. When *Y* learned that his remarks had offended Vaughn enough to cause her to consult a member of the corporation's legal department, he instructed *X* not to confront Vaughn about her job performance again and to take no action on any dissatisfaction he may have with her performance. *Y* also deliberately overstated his own satisfaction with Vaughn's performance, ensuring that she received only "satisfactory" performance ratings until she was terminated two years later during a reduction in force as one of the department's two lowest ranked contract analysts.

Vaughn claimed that her termination constituted disparate treatment prohibited by Title VII of the 1964 Civil Rights Act. To prevail, she must prove that her employer intentionally discriminated against her because of her race. Should it be salient to the outcome of Vaughn's case that *Y* testified he blocked *X*'s future criticism of Vaughn and inflated her performance ratings to avoid charges of discrimination against the corporation? Would it alter the outcome if he testified that he took those actions to insulate Vaughn against the prospect of discrimination by *X* or others, including against the prospect that *Y* himself had been unwittingly influenced by *X*'s criticisms? Would it alter the outcome if *Y* testifies that he forbade *X* from making further comments about Vaughn's socializing with coworkers and lack of productivity and decided to insulate her from negative performance evaluations because he came to believe that *X*'s concerns were unfair given Vaughn's recent return from maternity leave? In short, of what significance should *Y*'s motivations be to the disposition of Vaughn's case, and what burden must Vaughn bear in sorting through what may be multiple contemporaneous and contradictory motivations in order to prove intentional discrimination?

The disposition of Vaughn's case should turn on whether her employer intentionally treated her differently from other similarly-situated employees by denying

---

[120] The following hypothetical is adapted from the court's opinion in *Vaughn v. Edel*, 918 F.2d 517 (5th Cir. 1990).

her truthful and material information about her performance because of her race or sex.[121] This elegant solution yields the same outcome regardless which motivation is found to have guided the employment decision. Why then are there many who might find this solution elusive?

Two common misconceptions about the doctrine may have deterred readers from reaching this solution: that the intentionality requirement means that the employer must have acted with a discriminatory state of mind (that is, with the specific intention to harm or disadvantage Vaughn because of her race or sex) and that, even if the employer need not act with a *conscious* intent to harm, its actions are not culpable unless it can be said that they resulted from prejudice. I reject both misconceptions and argue that neither is required by the manner in which Title VII doctrine constructs intentional discrimination. However, I recognize that ambiguities in the doctrine and in the rhetoric used to express the doctrine have fueled these misconceptions, and so I turn now to an explanation of what I take these ambiguities to be.

B. *How intentionality avoids presupposing prejudice*

The statutory text of Title VII does not confine discrimination to any precise meaning, and nowhere does it restrict a finding of unlawful discrimination to situations in which the employer has acted intentionally. Rather, section 703(a) of the Act states that a employer commits an "unlawful employment practice" if it takes an adverse employment action against the plaintiff "because of such individual's race, color, religion, sex or national origin."[122] Of course, this phrase does not necessarily imply that the employer acted with any particular state of mind.[123] Nor would inferring a state of mind requirement be consistent with the overarching normative commitments of the statute. The Supreme Court has responded to the statute's ambiguity concerning the meaning of discrimination not by supplying its own definition but by constructing a series of doctrinal frameworks, including complex burdens of proof and production, intended to sharpen the inquiry regarding whether specific adverse employment actions were taken because of the plaintiff's protected class characteristic.

The Supreme Court began its development of disparate treatment doctrine in the 1973 decision *McDonnell Douglas v. Green*.[124] The plaintiff was an African-

---

121 This is indeed the basis for the circuit court's decision reversing judgment against Vaughn. 918 F.2d at 523.

122 42 U.S.C. §2000e-2(a)(1).

123 The statute refers to intent only where it provides under Section 706(g) that equitable relief and back-pay are available for unlawful conduct "intentionally engaged in" – construed to mean deliberately or not accidentally engaged in – and does not provide damages for harm caused for any reason other than discrimination. Finally, Section 1981A further distinguishes for remedial purposes between "intentional discrimination" and disparate impact where it provides for compensatory and punitive damages in relation to the former but not the latter.

124 411 U.S. 792 (1973).

American civil rights activist and mechanic who had been laid off by McDonnell Douglas purportedly pursuant to a reduction in force at one of the company's plants. Green protested the company's action through a "stall-in" organized to block other workers from entering the plan, and he also have been involved in one other demonstration at the plant – a "lock-in" – during which some of the defendant's employees were locked in the plant. Roughly a year following Green's layoff, McDonnell Douglas advertised for mechanics but denied Green's application to be rehired on the basis of his alleged participation in the in "stall-in" and "lock-in." Green brought suit for unlawful discrimination under Title VII on the basis of his race[125] and in retaliation for his protesting the company's purported race discrimination.[126] He had no direct evidence that the company had been motivated to select him for layoff because of his race.

In permitting Green's race discrimination claim to go forward on the basis of circumstantial evidence, the Supreme Court developed an elaborate burden-shifting scheme which it would continue to refine over the following three decades. The plaintiff bears the initial burden of demonstrating a prima facie case of discrimination, which Green satisfied by showing that he was black, that the defendant did not dispute his qualification for the position he had formally held, but despite his qualifications and the defendant's active solicitation of applicants he was denied rehire even though the position remained open.[127] The Court later explained that, in allowing the plaintiff to proceed without "direct proof of discrimination," the prima facie case "create[s] an inference that the decision was a discriminatory one."[128] The defendant may rebut this presumption by articulating a "legitimate, nondiscriminatory reason" for the adverse employment action.[129] The Court found that McDonnell Douglas satisfied this burden by providing evidence that it refused to rehire Green because of his participation in allegedly unlawful protests rather than because of his race, and as a result the Court concluded that the

---

125 Section 703(a) of the statute

126 Section 704(a) of the statute

127 The Court specified that a prima facie case of individual disparate treatment race discrimination will be satisfied by a showing:

> (i) that [the plaintiff] 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.

411 U.S. at 802. These elements are intended to remain flexible so that they can be adapted to a variety of adverse employment actions and factual contexts. *Id.* at 802 n.13.

128 *International Bhd. of Teamsters v. United States*, 431 U.S. 324, 358 n.44 (1977).

129 *McDonnell Douglas*, 411 U.S. at 802.

burden shifted back to the plaintiff to show that the defendant's proffered reason was a "pretext" for discrimination.[130]

The Supreme Court later explained that, under the *McDonnell Douglas* pretext analysis, the defendant's burden is merely one of production, not persuasion, and that "the ultimate burden of persuading the trier of fact that the defendant intentionally discriminated against the plaintiff remains at all times with the plaintiff."[131] Thereafter, the Court held that "the factfinder's disbelief of the reasons put forward by the defendant . . . may, together with the elements of the prima facie case, suffice to show intentional discrimination," but the rejection of those reasons does not compel judgment for the plaintiff.[132] Rather, the presumption of discrimination established by the prima facie case drops away once the defendant produces admissible evidence of a legitimate, nondiscriminatory reason and is not resurrected by disproof of that reason. Instead, the plaintiff must prove that the proffered reason was false, and "that discrimination was the real reason."[133]

It bears remarking that, when the Supreme Court discusses the *presumption* created by the prima facie case and the *reasons* proffered by the defendant, it makes no special effort to associate either with the defendant's state of mind. To the contrary, in the quoted passages, we see that the Court refers to a presumption of *discrimination* and a "real reason" that is *discrimination*. Of course, the term "discrimination" is a characterization of the act and not the motive, suggesting the that the Court has resisted equating intentional discrimination with proof of prejudicial motive but rather has equated it with the differential treatment that defines the act. Moreover, this conclusion is consistent with the normative commitments that the Supreme Court has identified as shaping its interpretation of Title VII and, specifically, the substantive protections of section 703.

The Supreme Court has recognized in section 703 a "basic objective of 'equality of employment opportunities' and the elimination of practices that tend to bring about 'stratified job environments.'"[134] The guiding commitment to equal treatment in

---

[130] *Id.* at 804-05.

[131] *See Texas Dep't of Community Affairs v. Burdine*, 450 U.S. 248, 254-55 (1981).

[132] *St. Mary's Honor Ctr. v. Hicks*, 509 U.S. 502, 511 (1993). Though the *Hicks* decision provoked considerable criticism from scholars who believed that it undermined by integrity of the McDonnell Douglas framework by weakening the prima facie presumption of discrimination and allowing untrue nondiscriminatory reasons to erode the very presumption by which the plaintiff was enabled to utilize circumstantial evidence, the Court later substantially allayed those concerns when it clarified that, indeed, there may be cases were the plaintiff need go no further than to disprove the proffered reason in order to sustain his claim. *See Reeves v. Sanderson Plumbing Prods., Inc.*, 530 U.S. 133 (2000).

[133] *Id*. at 510-11 & n.4 (internal quotation marks and citation omitted).

[134] *Burlington Northern & Santa Fe Ry. Co. v. White*, 548 U.S. 53, 63 (2006) (quoting *McDonnell Douglas v. Green*, 411 U.S. 792, 800 (1973)).

employment is balanced by the Court against the concern that the prohibition against discrimination may be used by plaintiffs or the government as a justification for encroaching upon the legitimate business decisions of private employers. As the Court has stated, "Congress has not commanded that the less qualified be preferred over the better qualified simply because of minority origins. Far from disparaging job qualifications as such, Congress has made such qualifications the controlling factor, so that race, religion, nationality, and sex become irrelevant."[135] Employers remain free to define legitimate criteria of employability, even where their choices prove inefficient, and they remain free to choose legitimate means to evaluate such criteria. The statute does not strip employers of the "discretion to choose among equally qualified candidates, provided that the decision is not based upon unlawful criteria."[136] What they may not do is apply such criteria unequally because of the plaintiff's social status. Thus, disparate treatment doctrine is constructed to permit the factfinder to differentiate between unequal treatment because of the plaintiff's status and legitimate exercises of business discretion concerning the employers' methods of employee evaluation, reward and discipline. The doctrine collapses these concerns into the evidentiary question whether the employer intentionally discriminated. Attaching a particular state of mind requirement would impair the doctrine's ability to balance these normative commitments and, in particular, it would artificially cabin the commitment to equal treatment.

Unfortunately, this conclusion remains ambiguous at the rhetorical level of Supreme Court opinions. For example, in dicta the Court has explained that disparate treatment "is the most easily understood type of discrimination. The employer simply treats some people less favorably than others because of their race, color, religion, sex or national origin."[137] This much of the Court's explanation is straightforward and confirms what its proof structure implies. Unfortunately, the explanation continues and in doing so becomes decidedly less clear, stating "[p]roof of discriminatory motive is critical, although it can sometimes be inferred from the mere fact of differences in treatment."[138] Certainly the Court's dicta leaves open to the plaintiff the opportunity to prove discrimination without direct evidence of prejudice. Depending on how the phrase "discriminatory motive" is to be understood, the Court may be taken to mean either that a "discriminatory motive" is nothing more than a motive to treat persons differently

---

[135] *Griggs v. Duke Power Co.*, 401 U.S. 424, 436 (1971). Although *Griggs* is a disparate impact case, for which no showing of intentionality is required, it interprets the same statutory provisions and the protection of employer discretion is at least as robust – arguably more so – in disparate treatment cases. *See, e.g.*, *Price Waterhouse v. Hopkins*, 490 U.S. 228, 243-44 (1989) (explaining disparate treatment doctrine to presume that, once the employer complied with the statute by ignoring the plaintiff's status, "it would naturally focus on the qualifications of the applicant or employee" and further clarifying that "no other qualification is affected by this title" (quoting Congressional record)).

[136] *Texas Dep't of Community Affairs v. Burdine*, 450 U.S. 248, 259 (1981) (disparate treatment case).

[137] *International Bhd. of Teamsters v. United States*, 431 U.S. 324, 335 n.15 (1977).

[138] *Id*.

because of their protected class characteristic or that differences in treatment are relevant to the extent that they support the conclusion that they arise due to prejudice.

However, to reach the latter conclusion about the Court's rhetoric would result in a false characterization of the doctrine and the normative commitments embedded in the intentionality requirement. The holdings of several disparate treatment cases make clear that intentionality is not to be taken as a synonym for conscious prejudice or animosity. Rather, the Supreme Court has repeatedly held that, where the plaintiff provides direct evidence of intentional discrimination (*e.g.*, in the form of a facially discriminatory policy), the plaintiff need not also prove animus. For example, in *City of Los Angeles, Dept. of Water and Power v. Manhart*,[139] the Court held that a policy requiring women to fund pension plan at a disproportionately higher rate than men based on plausible assumptions concerning life expectancy violated Title VII, because it did not pass the "simple test of whether evidence shows 'treatment of a person in a manner which but for that person's sex would be different.'"[140] Whether the defendant's true reasons for requiring women to pay a premium for pension coverage were rationally economical or prejudicial made no difference to the Court's decision. Similarly, in *Goodman v. Lukens Steel Co.*,[141] the Court held that a union's refusal to prosecute grievances of race discrimination asserted by its African-American members, though it otherwise prosecuted other grievances against the company, constituted intentional discrimination notwithstanding the absence of evidence of discriminatory animus.[142] As a final example, in *International Union, United Auto., Aerospace & Agr. Implement Workers of Am. v. Johnson Controls*,[143] the Court held that a policy prohibiting fertile

[139] 435 U.S. 702 (1978).

[140] *Id.* at 711 (citation omitted). *Accord E.E.O.C. v. Joe's Stone Crab, Inc.*, 220 F.3d 1263, 1283-84 (11th Cir. 2000) ("To prove the discriminatory intent necessary for a disparate treatment or pattern or practice claim, a plaintiff need not prove that a defendant harbored some special 'animus' or 'malice' towards the protected group to which she belongs."). *See also, e.g.*, *Trans World Airlines v. Thurston*, 469 U.S. 111, 121 (1985) (where "direct evidence that the method of transfer available to a disqualified captain depends on his age," under the Age Discrimination in Employment Act such a policy is discriminatory on its face).

[141] 482 U.S. 656 (1987).

[142] *Id*. at 667-68. In so holding, the Court decided that the plaintiffs' antidiscrimination rights against the union under § 703 of Title VII exceeded their rights to fair representation under the NLRA, which would have permitted the union to avoid liability if it could show that its reasons were not arbitrary – for example, that they were based on political calculations about the likelihood of success of dissimilar grievances, a desire to avoid antagonizing the employer or to satisfy and protect the interests of a majority of its members. The *Goodman* ruling precludes any of these reasons from serving as a basis for the union to avoid Title VII liability, even in the absence of evidence of animus, provided that the plaintiffs otherwise prove that the union refused to prosecute their claims because of race.

[143] 499 U.S. 187 (1991).

women, but not fertile men, from occupying certain hazardous positions violated Title VII, regardless of "[t]he beneficence of the employer's purpose."[144]

In sum, the intentionality requirement is not a two-stage requirement – that the plaintiff prove causation (*i.e.*, that the employer took adverse action against her because of her protected class characteristic) and then prove prejudice (*i.e.*, that the employer acted due to negative beliefs about or attitudes toward members of the plaintiff's class, or positive beliefs about or attitudes toward members of a class that benefited from the plaintiff's adverse treatment). Nor is it a requirement that the evidence of material causation presented by the plaintiff be limited to proof of prejudice (*i.e.*, that prejudice serve as the causal mechanism). Rather, evidence that an adverse action was undertaken because of the plaintiff's protected class characteristic may include evidence of rational[145] or attitudinally neutral, or even benevolent,[146] motivations. While prejudice may be instrumental in proving culpable causation in certain cases, proof of prejudice is not required where causation may be established by other means.

C. *How intentionality avoids presupposing conscious intent*

Returning now to Ms. Vaughn's case, the Fifth Circuit concluded that, "[i]n neither criticizing Vaughn when her work was unsatisfactory nor counseling her how to improve, [her employer] treated Vaughn differently than it did its other contract analysts because . . . she was black." The court of appeals did not rely on the *McDonnell Douglas* presumption to reach this conclusion. Rather, it based its conclusion on the magistrate judge's findings of fact, which determined that the employer "ignored its own procedures [regarding feedback and evaluation] for a racial reason."[147] The appellate court held that the disparate treatment to which Vaughn constituted intentional discrimination "however benign [the employer's] reason may initially appear to be."[148] Moreover, one must conclude from the court's opinion that the department head's use of the phrase "black matriarch" had no bearing on the court's decision and that, had he never expressed fear of litigation but instead, for example, expressed the desire to protect Vaughn against further harsh criticism by her supervisor, the ruling would have been no different. For Vaughn's employer would have subjected her to differential treatment, leading to an adverse employment action,[149] for a racial reason.

144

145 [*Manhart*]

146 [Johnson Controls?]

147 *Vaughn v. Edel*, 918 F.2d 517, 523 (1990).

148 *Id.* at 523.

149 Here, the court acknowledges that whether the employer's decisions ultimately damaged Vaughn's employment status "will never be known," for the department head's initial decision may have shielded her from earlier termination or other discipline. *Id*. at 523. But this does not change the fact that Vaughn would not have been terminated when she was without her employer's prior deviation from its normal feedback and evaluation procedures. Thus, the court further explained that it "will not sterilize a seemingly

But let's play out the hypothetical. What if the department head, *Y*, did not reach a conscious decision to appease Vaughn and avoid further controversy by providing her with satisfactory performance evaluations? What if, instead, *Y* relied on his own observations and the recommendations of supervisor *X* and, over the following 9 months, provided a series of negative performance evaluations culminating in her termination? Suppose that Vaughn brings a claim of disparate treatment discrimination. She introduces the "black matriarch" statement as direct evidence of intentional discrimination, because she believes that it demonstrates *Y*'s stereotypical belief, whether conscious or not, that African-American workers are lazy, excessively social and unprofessional. She further contends that these stereotypes clouded *Y's* impressions of her performance and that her negative evaluations were inaccurate because they were demonstrably lower than the evaluations of white contract analysts whose objective performance she either equaled or surpassed. *Y* testifies that, in providing the evaluations, he merely followed company procedure. Both sides dispute whether the evaluations were sufficiently informative to provide Vaughn with an opportunity to improve her performance, as well as whether her objective performance did in fact improve.

The present hypothetical obviously differs from Vaughn's actual case in several ways. One significant difference for our purposes is that, unlike the facts of the reported case, here *Y* is not conscious of having taken Vaughn's race into consideration. He believes that he applied the standards for evaluating employees in a race-neutral manner, and that Vaughn's evaluation fairly reflects her performance. Direct evidence of a causal connection between her race and either *Y*'s evaluations or her termination is now much weaker, making this hypothetical a more difficult case even though Vaughn now claims discrimination on the basis of a *negative* evaluation motivated by *Y*'s implicit prejudices (*i.e.*, his negative beliefs and attitudes toward her because of her race, and of which he is not aware) or implicit stereotypes (*i.e.*, his negative beliefs about Vaughn because of her race).[150] This version of the hypothetical is very much in line with the theory of implicit discrimination illustrated by the behavioral realists through their paradigm case.[151]

In Part II.B, we observed that the intentionality requirement is not a state of mind requirement restricting the plaintiff's proof of causation to a showing of prejudice, and we applied that lesson to Vaughn's case to show that disparate treatment for purportedly benevolent or neutral reasons satisfies the requirement of intentional discrimination. Now we consider the question whether the employer must be aware that

---

objective decision to fire an employee when earlier discriminatory decisions have infected it." *Id*. I am inclined to agree with those who may protest that the court did not settle an important causal puzzle here. But I must stress that the nature of the puzzle is not whether the employer's actions were taken because of race, but whether the actions that were taken because of race demonstrated adversity. [More on this point to come, though it will remain ancillary to the main thesis of the Essay.]

150 *See supra* note 1.

151 *See supra* note __ and accompanying text.

it has acted because of the plaintiff's protected class characteristic in order for its actions to constitute *intentional* discrimination.

That stereotyping is probative of intentional discrimination has been well-established since the Supreme Court's decision in *Price Waterhouse v. Hopkins.*[152] Ann Hopkins brought a claim of disparate treatment discrimination against her employer, the accounting firm Price Waterhouse, based on its refusal to admit her to partnership. Hopkins provided evidence that was denied partnership because she refused to conform to stereotypical gender roles of femininity (*e.g.*, by wearing make up and feminine attire) and passivity (*e.g.*, by avoiding aggressive or assertive behavior). The accounting firm claimed interpersonal skills and performance justified its decision, which it had made on the basis of comments and evaluations Hopkins solicited from the partnership. The Supreme Court declared that "we are beyond the day when an employer could evaluate employees by assuming or insisting that they matched the stereotype associated with their group."[153] The Court developed a mixed-motive analysis; as described by Justice Brennan writing for the plurality, that test provided that, if the plaintiff's protected class characteristic was a "motivating factor" in the employer's adverse decision, then the employer was liable for discrimination even if it also, simultaneously, harbored other reasons for its action.[154]

The *Price Waterhouse* plurality declined to limit the types of stereotyping evidence that might be probative of discriminatory intent to the type of role-conformity evidence discussed in that case.[155] The facts of the case make clear, however, that the

---

[152] 490 U.S. 228 (1989). The *Price Waterhouse* case is not alone; the Supreme Court has repeatedly affirmed that employers commit disparate treatment discrimination by subjecting members of a particular status group differently because of status-based stereotypes, without requiring plaintiffs to show that the stereotypes were consciously held. *See, e.g.*, *Manhart*, 435 U.S. at 707 n.13 (stating that Title VII prohibits 'the entire spectrum of disparate treatment of men and women resulting from sex stereotypes"); *Hazen Paper Co. v. Biggins*, 507 U.S. 604, 610-11 (holding that employment actions "based in large part on stereotypes unsupported by objective fact" represent the "essence of what Congress sought to prohibit in the ADEA"). [I note that *Manhart* and *Hazen Paper* show some tension concerning whether a stereotype must be inaccurate in order to be probative of discrimination, but in general whether an action caused by consideration of a statutorily protected characteristic is rational or efficient not be considered to shield the defendant from liability. More to come on this point.]

[153] *Id.* at 256.

[154] [Future draft will expand analysis, but for now this will suffice as Congress has amended the test to reconcile differences between the plurality opinion and Justice O'Connor's concurrence.]

[155] Following the relevant cognitive psychology research, we may call this form of stereotyping "prescriptive" in contrast to "descriptive" stereotyping whereby a person filters their observations and evaluations of an individual through beliefs that they hold about members of the individual's social group. *See* Diana Burgess & Eugene Borgida, *Who Women Are, Who Women Should Be: Descriptive and Prescriptive Gender Stereotyping in Sex Discrimination*, 5 Psych., Pub. Pol., & L. 665 (1999). By

Court included both conscious stereotypes (evidenced by Price Waterhouse partners who commented that Hopkins should dress more femininely, wear make-up, and take lessons at charm school)[156] and unconscious stereotypes (evidenced by partners who criticized Hopkins for being too aggressive, though they otherwise revealed no awareness that they had penalized Hopkins for her aggressiveness because of her sex).[157]

In codifying a revised version of the mixed-motive test in the 1991 Civil Rights Act, Congress declined to provide any further limitation on probative evidence. Rather, it simply provided that a showing that the plaintiff's protected status was a "motivating factor for an employment practice" will suffice to prove an unlawful employment practice, "even though other factors also motivated the practice."[158] In practice, many circuits continued to follow the instruction provided by Justice O'Connor in her *Price Waterhouse* concurrence, where she concluded that mixed-motive analysis should be available only where direct evidence of discriminatory motive is shown.

The Supreme Court's recent decision in *Desert Palace, Inc. v. Costa*[159] has placed disparate treatment doctrine at a crossroads, by expanding the use of mixed-motive analysis to claims relying on circumstantial evidence. Though this more expansive application of the motivating factor test introduces intriguing possibilities for implicit bias evidence, the *Desert Palace* opinion expressly reaffirms the intent requirement and otherwise provides no further interpretation of either intent or motive.

For the most part, stereotyping cases in the lower courts concern role conformity – that is, whether the victim was discriminated against for failure to conform to her stereotypical role.[160] This is the kind of prescriptive stereotyping addressed by the Court in *Price Waterhouse*. Conversely, cases in which implicit bias is alleged to have poisoned the assumed neutrality of the employer's decision-making process – e.g., by overemphasizing negative characteristics and undervaluing positive characteristics of

---

"prescriptive stereotype," I mean a belief about how a member of a group should behave based on her group membership. *Id.* at __. Psychologists writing the *amicus* brief on behalf of Hopkins equated prescriptive stereotypes with normative beliefs specifying "behaviors that are thought to be not only characteristic of each sex, but also desirable and encouraged." [Brief for the APA] at 1065; *see also* Fiske, *Controlling Other People*, [full cite] at 623 (arguing that the penalties for non-conformity to prescriptive stereotypes are "swift and severe").

156 [Cite to *Price Waterhouse*.]

157 [Cite to *Price Waterhouse*.] It is probably equally the true that the facts of the case reveal both benevolent and hostile stereotypes, attributing to women both positive and negative attributes and then forcing Hopkins to navigate those preconceptions, few of which she appears to have conformed to. E.g., Price Waterhouse, __. See generally Peter Glick & Susan T. Fiske, *The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism*, 70 J. Personality & Soc. Psych. 491 (1996).

158 42 U.S.C. § 2000e-2(m).

159 539 U.S. 90 (2003).

160 [Cite cases]

women or minority persons – are much more rare and typically less successful unless the plaintiff can show that a change in evaluation coincided with a change in the identity of the person responsible for performing the evaluations.[161]

This is the kind of case raised by our hypothetical version of Ms. Vaughn's case, but her case does not involve a sudden change of supervisor. How must Vaughn go about proving her case? In one of two ways. Either Vaughn must show that *Y*'s proffered reason for her termination (*i.e.*, that he properly evaluated her performance on a race-neutral basis and fairly concluded that her performance was wanting), though an honest account of his understanding, is not the true reason for her termination because *Y* was unknowingly influenced by his implicit bias against Vaughn as an African-American woman. (This is the type of showing for which the behavioral realists rightly argue that implicit bias research may prove useful.) Alternatively, Vaughn may demonstrate that *Y* has mischaracterized not his *reason* but his *action*. For example, Vaughn may demonstrate that white employees receiving the same evaluation scores that she received were not terminated. (This option is foreclosed by the facts of this case, but remains available to plaintiffs generally as a means of proving status-based disparate treatment.) She may concede that she was fired because of her evaluation, but otherwise prove that *Y* did not conduct her evaluation in the neutral and fair manner that he asserts. She may do this by circumstantial evidence showing, for example, that *Y* ranked white employees with objectively comparable or inferior performance higher than he ranked her. On the basis of this showing, the factfinder may infer that Vaughn was subjected to disparate treatment because of her race.

Moreover, disparate treatment doctrine does not require that, in making that inference, the factfinder also conclude that *Y*'s evaluations were motivated by prejudice.[162] In other words, disparate treatment because of race need not be construed as evidence of disparate treatment motivated by prejudice in order for it constitute actionable discrimination under the statute.

Finally, disparate treatment doctrine counts as evidence of *intentional* discrimination Vaughn's proof either that *Y*'s actions were motivated by implicit bias (such that his "honest" reasons were mistaken when compared with his true reasons) or that mischaracterized his own action (e.g., as executed in a neutral and fair manner). The supervisor's action did not cease to be intentional just because he was mistaken about either his motivation or his description of the action. If the behavioral realists believe as they claim that the intentionality requirement is incompatible with such cases of implicit discrimination because, for his action to be intentional, the discriminator must be "consciously aware, at the moment of decision, that he or she is discriminating" then it must not count either of Vaughn's proofs as evidence of intentional discrimination, even though current disparate treatment doctrine does so. This seems an odd result, and one issuing from the behavioral realists failure to interrogate what the doctrine means by "intentional discrimination" in a more robust and searching way. Do the behavioral

---

161 *E.g.*, *Thomas v. Eastman Kodak Co*., 183 F.3d 38 (1st Cir. 1999).

162 *See supra* Part II.B.

realists believe that any concept of intentional discrimination that does not claim *conscious intent to discriminate* as a key component is analytically incoherent? Do they believe that such an understanding of intentional discrimination would be so contrary the views ordinarily expressed by laypersons regarding the relationship between attributions of intentionality, or conscious motivation, and judgments of culpability as to be of no use to a factfinder?

In Part III, I aim to demonstrate that the doctrine's broad formulation of intentional discrimination is neither irrational nor impractical. It is exactly what the behavioral realists would need it to be in order to make use of implicit bias research without substantially remaking the doctrine; and, therefore, their critique of the requirement fails. Better would have been the development of an understanding of the sorts of intentional discrimination that implicit bias research can help to us rationalize – that is, an understanding of the sorts of intentional discrimination for which a person's unconscious motivations may count as reasons for their actions.

III. Why the critique of intentional discrimination fails

The behavioral realist account provides an inaccurate and excessively narrow interpretation of the intentionality requirement as a prelude to jettisoning that requirement as incompatible with a certain "truth" about discrimination – namely, that discrimination is often caused by implicit bias rather than an intent to act on conscious animus. But it is not necessary to jettison the intentionality requirement in order to target discrimination caused by implicit bias. The intentionality requirement does not require proof of a particular state of mind such as hostility or animus,[163] and it does not require conscious awareness by the employer of its discriminatory reasons for acting.[164] Understood in this way, the intentionality requirement is broad enough to consider evidence of an agent's unconscious motivations as reasons for his actions. Moreover, there is no inherent conflict between the requirement that disparate treatment discrimination be "intentional" and the notion that discriminatory behavior may be motivated by implicit bias. An agent may mistake a reason that he may have had to act for the true cause of his action, but this does not render his action unintentional.[165]

---

163 *See supra* Part II.B.

164 *See infra* Part II.C.

165 [**Try to move into main text.**] In this sense, the rationalist approach treated as a foil by the behavioral realists overreaches with respect to placing importance on the agent's conscious intentions. It is not because we take the agent to be the most objective or knowing observer of his reasons for acting that we are interested in the agent's statement of his intentions. An agent's intentions matter as an important factor in the classification of his actions – particularly in ascribing meaning to his actions – but we do not ordinarily credit the agent's expression of his intentions with the power to undermine our assessment that his actions were volitional, that they were subject to his control and no one or nothing else's.

I defend the intentionality requirement in further detail below.

First, the critique of intentional discrimination depends upon an excessively narrow account of intentionality, fundamentally at odds with Supreme Court doctrine.

Notwithstanding their bold rejection of the normative critique of their project,[166] Krieger and Fiske acknowledge (as they must) that the law's "normative goals" must be specified before the operationalization of those goals may be empirically tested. To that end, Krieger and Fiske never fully define just how much of the empirical model of implicit bias is relevant to the normative account of a claim of disparate treatment discrimination. Nor, for example, do they explain why they treat the concept of "intentional discrimination" as a descriptive error – a false empirical premise within the doctrine – rather than as a normative position around which the doctrine is built (i.e., that employers ought to be liable for, and courts ought to address with particular remedies, "intentional" discriminatory acts rather than unintentional acts or accidents).

Critics of the intentionality requirement typically do not provide a concise description of what they believe courts, including the Supreme Court, mean by "intentional discrimination" or "discriminatory intent." Krieger and Fiske are no different. In explaining their objection to the intentionality requirement, they purport to rely on a definition of intentional discrimination provided by Justice Brennan in his plurality opinion in *Price Waterhouse v. Hopkins*. There, Justice Brennan explains the plurality's construction of the motivating factor test,[167] stating that:

> In saying that gender played a motivating part in an employment decision, we mean that, if we asked the employer at the moment of the decision what its reasons were and if we received a truthful response, one of those reasons would be that the applicant or employee was a woman.[168]

We may say in this simple case that the agent suffers from self-deception. The agent may even suffer from weakness of will if the agent undertook to act on the basis of one set of goals and commitments (e.g., to provide a fair and unbiased evaluation of job performance, which if he had done so would have caused him to rank a female employee above a male employee) but then acted according to other unconscious attitudes (e.g., by undervaluing the superior performance of the female employee and overvaluing the performance of the male as a basis for awarding the latter a promotion). But in either case we do not say that, for example, the promotion or the preceding formal evaluation were unintentional, though we may say that the agent did not intend to distort the evaluations that he intentionally performed.

[166] See supra note __ and accompanying text.

[167] *See supra* Part II.3.

[168] 490 U.S. at 250, *quoted in* Krieger & Fiske, *Behavioral Realism in Employment Discrimination Law*, *supra* note **Error! Bookmark not defined.**, at 1010.

According to Krieger and Fiske, this passage "reflects two 'common sense' theories about the nature of discriminatory motivation,"[169] and they are furthermore clear that they view these theories rather as assumptions proven to be false by cognitive psychologists working in the field of implicit bias. These assumptions are (i) that the discriminator is "consciously aware, at the moment of decision, that he or she is discriminating" (also referenced as "transparent mental processing")[170] and that (ii) perception may be divorced from decision making, such that when the agent undertakes to render an "employment decision" he or she is capable of setting aside whatever defects or biases of perception preceded that decision.[171] Whether these assumptions are false is of no moment for this Essay,[172] because the more pressing matter is whether they are assumptions that drive the doctrine or even properly describe Justice Brennan's statements.

In Part II.C, we saw that conscious awareness of a discriminatory motivation is not required by the doctrine. Indeed, the holding of *Price Waterhouse* and the facts on which that case relies command this conclusion.[173] Krieger and Fiske are simply wrong to have taken Brennan literally here. His passage is more of a thought experiment than a description of reality. He does not mean that there would necessarily ever be a time when a plaintiff or a court could "*at the moment of*" decision ask the employer why it took that decision and receive an honest, inculpatory answer.[174] Nor does Justice Brennan mean to suggest that only evidence memorializing such an answer will show that the employer was motivated by the plaintiff's sex, since it is clear that the Court considers both evidence of comments submitted to the committee reviewing Hopkins promotion to partnership and statements made outside the committee review process about Hopkins generally.[175]

In sum, Krieger's and Fiske's criticism of the intentionality requirement says more about their commitment to addressing a particular form of discrimination – *implicit discrimination* (i.e., in which the defendant honestly professes a nondiscriminatory reason and yet is unwittingly motivated by implicit bias, as illustrated

---

[169] Krieger & Fiske, *Behavioral Realism in Employment Discrimination Law*, *supra* note __, at 1010.

[170] Id.

[171] Id.

[172] [In fact, I am inclined to agree that they are.]

[173] See supra note __ and accompanying text.

[174] Perhaps Krieger and Fiske are misled here by the flat tone of Justice Brennan's passage. Surely, they would not make this mistake if the Justice had written that gender will play a motivating part in an employment decision, "if, when we could lift the employer's thinking cap at the moment of decision, we would see that the employer was thinking about the plaintiff's gender." Then, the passage would be revealed to be a heuristic device and mere thought experiment, which is exactly what it is.

[175] [Cite to *Price Waterhouse*, or make this point clear in Part II.3 and cross-reference.]

by the paradigm case)[176] – than it says about disparate treatment doctrine as the Supreme Court has described it. Krieger and Fiske articulate an excessively narrow view of intentional discrimination, in which the discriminator holds certain prejudicial beliefs or attitudes against members of a protected class which he consciously takes to be reasons for committing actions that are adverse to members of that class and he intentionally undertakes such actions in order to fulfill his reasons.[177] Krieger and others have further argued that this concept of intentional discrimination presumes that persons who commit discriminatory acts have a disposition to do so – a "taste for discrimination" to borrow from Gary Becker[178] – that overrides other rational impulses and causes them to act according to their prejudices. Krieger and others have then used cognitive psychology in order to undermine each of these assumptions. But in doing to they have failed to reckon with a more expansive and formidable construction of intentional discrimination.

The concept of intentional discrimination does not require any particular disposition or state of mind. Rather, Krieger and Fiske have collapsed an important distinction between intentions and intentional actions. A person may *act intentionally* (i.e., in an intentional manner) and yet be motivated by beliefs or attitudes of which he is not aware. An agent may be mistaken with regard to his reason for acting, but this does not render his act unintentional. We need not define acts of intentional discrimination by revealing the agent's prejudicial intentions or motivations. Only insofar as the latter are relevant to reveal an illicit cause – that the action was taken because of the plaintiff's social status. Once we recognize this, we see that there is no conflict between the requirement that an adverse action intentionally discriminate against protected class members (*i.e.*, that it intentionally treat them differently from similarly situated nonmembers) and the observation that a significant amount of discriminatory behavior is motivated by implicit, or unconscious, attitudes and beliefs. In fact, we find a new opportunity to think more clearly about what we mean by intentional discrimination and how we understand their relation to unconscious motivations. In Part III.B., I will begin to sketch out a path by which we may begin to do this.

Second, the critique of intentional discrimination overlooks the normative significance of the intentionality requirement. As discussed in Part II.B., the concept of intentional discrimination aims to balance a robust commitment to equal treatment in the workplace with a commitment to avoid encroaching upon legitimate exercises of employer discretion. As the Justice Brennan states on behalf of the *Price Waterhouse v. Hopkins* plurality, disparate treatment doctrine assumes that once the employer complied with the statute by ignoring protected characteristics (e.g., race, sex, national origin, religion), "it would naturally focus on the qualifications of the applicant or employee."[179]

---

[176] *See supra* Part II.B.

[177] See supra note __ and accompanying text (discussing the "simple view of intentional discrimination").

[178] [CITES]

[179] *Price Waterhouse v. Hopkins*, 490 U.S. 228, 243-44 (1989). In fact, Congress has made clear that "no other qualification is affected by this title." (*Id*. at 244 (quoting Congressional record)).

Thus, in the very same opinion where Krieger and Fiske identify the rational transparency and perception-judgment assumptions, the Supreme Court also reaffirms the statute's normative commitment to preserve employer discretion by differentiating between intentional discrimination and legitimate exercises of employer discretion.

The behavioral realists account should be able to address this commitment as it answers other basic normative questions. Such as, what sorts of conduct is the intentionality requirement meant to exclude? Conduct involving equal application of legitimate evaluative criteria, uninfluenced by the plaintiff's social status, regardless whether the conduct resulted in harm to the plaintiff.[180] And, much more unlikely to be the basis for an actual adverse employment action, accidental conduct resulting in a difference in treatment of employees of one social status as compare with employees of another status.[181] Is it wrong for the doctrine to exclude these types of conduct from liability, to balance the normative commitments of equality and employer discretion in this way? Again, it is not intended to, nor does it exclude, conduct for which the employer is mistaken about its reasons, and so exclusion of the behavioral realists' paradigm case cannot be a basis for rejecting the requirement.

Krieger and Fiske argue that we should substitute for intentional discrimination a showing of "discriminatory motivation," including evidence of any implicit bias that give rise to discriminatory conduct. But, as I have explained, the intentionality requirement does not exclude intentional discrimination against protected class members just because the discrimination occurred because of implicit bias. Rather, the intentionality requirement ensures that defendants are held liable for discriminatory conduct over which they may exercise a reasonable degree of control, as opposed to their accidents. An employer is not asked to account for the thoughts of its agents[182] but for unequal treatment of the plaintiff over which the employer exercises reasonable control when that treatment (not simply the harm) is caused by the plaintiff's social status. Wholesale adoption of implicit bias research as a normative template for the restructuring of employment discrimination law suggests that employers should be held liable for behaviors that may not be subject to corrective control because the attitudes that motivate them are not available for conscious evaluation and because (as Krieger and Fiske frequently remind their readers) perception and judgment are not separable mental events. Such an approach does not adequately explain the basis for employer liability because it does not help us to distinguish between prohibited unequal treatment and legitimate exercises of employer discretion; rather, it suggests that often no such distinction can be

---

[180] This sort of policy or practice could, however, form the basis of a disparate impact challenge.

[181] The nature of workplace decisions such as hiring, firing, or other employment actions that an employer may take with respect to an employee make it difficult to imagine such a case, and so it is difficult to know whether we must take this formulation seriously. For further development of this point, see infra Part III.B.

[182] Though prophylactic measures against certain forms of implicit bias may be prudent and are consistent with restraining intentional discrimination.

made.[183] The very notion of the legitimacy of evaluative criteria is placed under constant threat from the assumption that implicit bias, rather than legitimate business reasons, may have driven the choice of criteria.

Moreover, the proposed "discriminatory motivation" requirement, while it may function in some cases to facilitate proof of a causal connection between the plaintiff's social status and unequal treatment, may also drive the factfinder away from robust consideration of proof of unequal treatment just as it shifts the factifinder's attention toward the defendant's mental state. The intentionality requirement authorizes a limited inquiry into those choices made by the employer that result in adverse action against the plaintiff and the causes of those choices, but it does not authorize an all-out investigation of the rationality or efficiency of employment decisions were authorized by an inquiry into the defendant's explicit and implicit mental states. Any serviceable adaptation of implicit bias research that would be compatible with the normative commitments of disparate treatment doctrine will have to have to accept some distinction between perception and judgment (*e.g*, between bias – even in an evaluation process – and action on the basis of that bias); it will have to differentiate between the *activation* of implicit negative beliefs and attitudes (the most cognitively automatic dimension of implicit attitudes and stereotypes) and the *application* of such beliefs and attitudes to particular employment actions (the area where the prospects for conscious control arguably are greatest).

Third, the critique of intentional discrimination, when it also advocates wholesale adoption of implicit bias research, suffers from two critical flaws in its overestimation of that research. The first flaw is that it overestimates the conclusion that implicit biases (and the behaviors to which they may be causally link) develop and spring forth automatically. Increasingly, cognitive psychologists have shown that implicit biases are "malleable" in that they are subject to conditions such as motivated reasoning and cognitive attention (*e.g.*, to the possibility of discrimination or to social norms against discrimination).[184] Also, psychologists have come to emphasize the distinction between activation of implicit biases and their application in the context of social behaviors. Awareness of behavior may be greater than awareness of motivation, depending on the type of behavior – and this brings us to the second flaw in the overestimation of implicit bias research: Legal scholars, including Krieger, have overestimated the link between implicit attitudes and legally salient behaviors, because they have collapsed the distinction between the sorts of nonverbal behaviors most often observed in psychological experiments and the sorts of deliberative (even collective) behaviors that give rise to a finding of discriminatory conduct in the workplace. Krieger is correct that discriminatory intent serves as a kind of restrictive indicator of causation in disparate

---

[183] [Or suggest further normative command of suppression of implicit bias, including through intentional control.]

[184] *See, e.g.*, Irene V. Blair, *The Malleability of Automatic Stereotypes and Prejudice*, 6 Personality & Soc. Psych. Rev. 242 (2002); Ziva Kunda & Lisa Sinclair, *Motivated Reasoning with Stereotypes: Activation, Application, and Inhibition*, 10 Psych. Inquiry 12 (1999).

treatment cases (and I agree that the absence of conscious intent should not preclude a finding that a person's protected class status caused her disparate treatment, including due to the employer's implicit biases), but implicit bias cannot serve as an adequate substitute for proof of causation if it can only be shown to be a cause of behaviors that would not in themselves rise to the level of an adverse employment action.[185] Instead, there must be a evidence from which a reasonable factfinder could infer a connection between an implicit bias and an intentional action.

The second flaw concerning Krieger and Fiske's overestimation of the implicit bias research (discussed in further detail in Part IV.A.) is that cognitive psychologists have not been able to provide a clear demarcation regarding which implicit biases are truly "negative," and therefore a proper basis to attribute "discriminatory motivation" to the employer and which are merely benevolent or benign. Because the behavioral realists treat implicit biases as though they were implicit prejudices – because the notion of a negative implicit belief or attitude is critical to the type of motivation they seek to prove using implicit bias research – they never truly rid themselves of the question of when a bias regarding the plaintiff is a truly discriminatory bias in the sense that it is negative, and when it is something else.

Fourth, and finally, the critique of intentional discrimination overlooks the intuitive appeal of the requirement of intentionality as a reason to assess blameworthiness. Significant psychological research and work in experimental philosophy demonstrates that lay persons (including jurors as well as judges) use attributions of intentionality as a means to assess blameworthiness. Asking a person to hold a defendant responsible without attributing intentionality to the defendant makes the reason for assigning responsibility harder to identify, weakening the intuitive appeal of the plaintiff's claim. Why do this if it is not necessary to preserve space to include unconscious motivations as part of an explanation of the defendant's actions? There is no reason to abandon the intentionality requirement if it is broad enough to permit plaintiffs to sustain claims based on defendant's unconscious motivations, even if the defendant believed it had acted for legitimate, nondiscriminatory reasons.

Krieger and Fiske resist characterizing discrimination based on implicit bias as dispositional in nature. Whether or not their effort to avoid rearticulating disparate treatment as discrimination caused by a prejudicial *disposition* is successful, Krieger and Fiske evidently wish to expand the scope of what we might otherwise call "prejudice" to include implicit beliefs and attitudes, even those that do not manifest hostility but only some other form of negativity, such as aversion. They appear to feel the need to preserve an aspect of the doctrine that they have otherwise complained is empirically inaccurate (at least, in its simplest form) just because it provides a compelling narrative for assessing culpability. That is, they appear to wish to preserve the notion that a person discriminates when they act according to illicit, negative motivations – when they are disposed to act out of prejudice – even though a more inclusive form of

---

[185] Though behaviors such as reluctance to engage in conversation, avoidance of eye contact and other nonverbal forms of aversion may be germane to a claim of hostile work environment harassment.

antidiscrimination protection would ignore the defendant's motivation altogether unless it were necessary to prove that disparate treatment was caused by the plaintiff's protected class status. Thus, it seems that behavioral realists view prejudice and not intentionality as the crucial component for assessing blameworthiness for discriminatory acts. But this assumption contradicts significant psychological research.

Where is the realism of behavioral realism with respect to the role ordinarily played by attributions of intent in lay judgments of culpability? Ironically, most behavioral realists have not sought to combine their *realistic* account of discrimination (as provided by implicit bias research) with a similarly realistic account of lay attributions of intent or judgments of culpability. If intentionality has a descriptive more than a normative function, as behavioral realists such as Krieger and Fiske contend, perhaps that function exists to mirror the ordinary, common sense framework by which lay persons who may be called to serve as jurors make determinations of intentionality and culpability. This concern is discussed by Gary Blasi[186] as a reason why, at its outer limit, a civil liability regime cannot incorporate enough of the theory of implicit discrimination into its framework to address a broad swathe of real-world phenomena. Instead, Blasi takes this as a reason why the law should turn to a regulatory regime that facilitates the development of "best practices" in combating discrimination motivated by implicit bias.[187] I reject the pessimism of this approach but applaud Blasi's concern for what Mark Alicke has called the "psychology of blame."[188]

According to Alicke, the blameworthiness of an action turns on the degree of "culpable control" exercised by the actor. The culpable control model comprises a volitional control aspect (which concerns the degree of volitional control exercised over the action and its outcome), a mental aspect (which concerns the relationship between the actor's mental processes – including plans, motives, desires – and the action), and a behavioral aspect (which concerns the nature of the behavior as an action or omission).[189] Significantly, Alicke's model is a flexible one in which neither the first nor second aspect is interpreted to require that the actor form a specific intention to act in order to be judged culpable. Rather, Alicke notes the importance of "motives" in addition to intentional plans, and he states that the primary issue relating to the control aspect is whether the behavior is volitional or accidental. In fact, contrary to the pessimistic import of Blasi's interpretation, Alicke in fact argues that, once charged to judge blameworthiness, subject engage in a "blame-validation mode of structural linkage assessement" whereby they review evidence "in a biased manner by exaggerating the actor's volitional or causal control" and even "lowering their evidentiary standards for blame."[190] Laypersons do not require transparent information about the subjects intentions or dispositions, and the

---

186 And by him alone among behavioral realists so far as I have been able to determine.

187 See Blasi, supra note __, at __.

188 Mark D. Alicke. *Culpable Control and the Psychology of Blame*, 126 Psych. Bulletin 556 (2000).

189 *Id.* at 557-59.

190 *Id.* at 558.

absence of such evidence does not necessarily undermine the strength of their convictions about moral responsibility.

If Alicke is correct in describing lay constructions of blameworthiness in this manner, his theory supports the broad conception of intentionality endorsed by disparate treatment doctrine as having the further benefit of providing a psychological anchor for lay factfinders to catch hold of in order to steady their resolve to make findings of liability where the evidence lacks clear evidence of conscious intent. In sum, laypersons take a flexible approach to the question of what defines an intentional action when assessing culpability, and this approach facilitates judgments of culpability. But we take away the advantages of this flexible approach if we say that intentionality is not germane to a judgment of discrimination, and we consequently risk provoking the factfinder to compensate by applying other elements of liability more restrictively. Admittedly, there may be reasons to be wary of authorizing or empowering factfinders to pursue their blame-validation strategies unbounded, and the very existence of a blame-validation framework for information processing may in some cases undermine our confidence in the correctness of judgments of discrimination. But Alicke in particular demonstrates that the notion of volitional control that is salient for lay determinations of blameworthiness has more in common with broad conception of intentionality turning on the manner of action discussed above than the rationalists' simple view of intentional discrimination.[191] My point here is that behavioral realists could potentially benefit from examining the intentionality requirement under the same methodological lens of cognitive psychology as they have used to discover the causes and modalities of discrimination generally, and were they to do so they may be encouraged to endorse a more flexible understanding of how we make judgments of intentionality.[192]

One further area of research reveals that laypersons may seek relief from the responsibility to link their judgment that an action is discriminatory with an evaluation of the disposition or state of mind of the actor. Janet Swim and her colleagues have demonstrated through experiments that test subjects are capable of judging conduct discriminatory without clear evidence of prejudice or discriminatory intent, and when evidence of intent is either absent or ambiguous subjects often detach attributions of prejudice (which principally turn on evidence of intent) and judgments of discrimination.[193] Put another way, we should not expect consistency between attributions of prejudice (which are personal to the agent whose actions are under

---

191 See supra Part I.B.

192 I understand that the behavioral realists purport to be describing what the court's have done with the concept of intentional discrimination in making their initial diagnosis of its limitations. But in concluding that the intentionality requirement should be jettisoned, they move from diagnosing a problem with the doctrine – on which I argue does not exist, at least with regard to Supreme Court doctrine – to speculating that intentionality is incompatible with the view that discrimination may result from implicit bias regardless how we define *intentional action*.

193 See Janet K. Swim et al., *The Role of Intent and Harm in Judgments of Prejudice and Discrimination*, 84 J. Personality & Soc. Psych. 944 (2003).

evaluation) and judgments of discrimination (which concern the nature of the action under evaluation and the harm that it causes, not just whether the relation of action to injury were part of the agent's intentional plan) such that the latter requires the former; lay persons simply do not impose such a requirement on their own determinations.[194] In fact, because attributions of prejudice are understood to be judgments about the character of a person whereas judgments that a person's actions are discriminatory are not, laypersons may seek ways to express moral disdain for a person's actions without impugning the person's character in any enduring sense that speaks to a person's dispositions. As Swim and her colleagues have noted, "people may be more confident about labeling a particular behavior as discriminatory than generalizing from one behavior to the character of an actor, an attribute that may be presumed to have cross-situational consistency."[195]

IV. <u>Repudiating prejudice as a requirement of intentional discrimination</u>

A. *The instability of the concept of prejudice*

There is something imprudent in recommending that disparate treatment doctrine abandon the intuitive appeal and normative balance of the intentionality requirement in order to make more robust use of evidence of implicit bias, given that what constitutes an implicit bias is an inherently unstable idea. To what degree should the probative value of evidence of implicit bias turn on the extent to which the identified bias represents an attitude or belief regarding the plaintiff that is *negative*? Of course, under my view, prejudice is not absolutely required. But this does not answer the question whether under Krieger's and Fiske's account of discriminatory motivation some notion of prejudice is retained just insofar as to be relevant a motivation must constitute a negative bias against the plaintiff.

Reviewing implicit bias literature, Hal Arkes and Philip Tetlock protest that "the case has not yet been made that implicit prejudice is prejudice."[196] Looking primarily at the use of the term "prejudice" in political and sociological discourse, Arkes and Tetlock argue that prejudice is a "value-laden" judgment that is not necessarily represented by the negative implicit attitudes identified by other researchers to be components of implicit prejudice.[197] Drawing upon the definition provided by Gordon

---

[194] *Id.* at 945.

[195] *Id.* Swim and her colleagues rightly acknowledge that this tendency may be reversed in special cases where factfinders understand the agent's actions to be constrained such that he is prevented from treating persons differently in a way that could be considered discriminatory. In such as special case, they acknowledge that "[i]f a person knows tha tan actor wants to discriminate against a woman" but is precluded from doing so, then the person may be more likely to judge the actor to be prejudiced than the actor's behavior to be discriminatory." *Id.* at 957.

[196] Arkes & Tetlock, *supra* note __, at 258.

[197] *Id.*

Allport, Arkes and Tetlock deduce that prejudice as a negative attitude must be "functionally intertwined" with beliefs that "indiscriminately" attribute negative qualities to group members; that it must reflect animus or hostility (rather than other forms of aversion such as guilt, anxiety or avoidance); that it must be unwarranted and resistant to change; and that "it must be truly negative *as opposed to less positive* than the affect one has toward other groups."[198]

The problem that this account poses for the behavioral realists is not that it is descriptively accurate[199] but that, wherever the line is to be drawn between implicit biases that are or are not prejudice, they have been unable to draw it. And if they will not draw such a line, then they must be ready to defend the view that even benign and benevolent beliefs and attitudes may be relevant to an investigation of implicit discrimination if they are relevant to show that disparate treatment occurred because of the plaintiff's protected class status. And they must do this without the argument that any ascriptive belief or general attitude about a group is negative, since that would equate even the most dehumanizing of ascriptive beliefs with feelings of shame and guilt that a true egalitarian would have in contemplating his own undeserved privilege when in the presence of someone who lacks similar advantages.

Of course, even assuming that Arkes and Tetlock have provided an accurate account of the *traditional* understanding of prejudice, it is overly restrictive and incomplete as a normative basis for conferring liability for discrimination.[200] Antidiscrimination law does not excuse disparate treatment motivated by prejudice just because the perpetrator targets some group members, but not others, based on their status or because the perpetrator intended to act on more positive or favorable attitudes toward ingroup members as opposed to "truly negative" attitudes toward outgroup members.[201] The law does not require the reasons for which the perpetrator treats some differently than others to be inaccurate or misinformed.[202] It does not require that the perpetrator's

---

198 *Id.* (emphasis added). *See also* Philip E. Tetlock & Hal R. Arkes, *The Implicit Prejudice Exchange: Islands of Consensus in a Sea of Controversy*, 15 Psych. Inquiry 311 (2002).

199 Samuel Bagenstos has rightly shown that this is not a criticism based on the scientific falsity of implicit bias theory, see Bagenstos, Implicit Bias, "Science," and Antidiscrimination Law, supra, at __. Furthermore, the brief history of prejudice studies provided in Part I.A. shows this criticism to be anachronistic. Clearly, Banaji and her colleagues are right to respond that psychologists have rejected such limitations on the definition of prejudice for some time. See Banaji et al., No Place for Nostalgia, supra, at __.

200 See infra __. It is hardly apparent to most in our day and age why prejudice must be akin to hatred or why group-based hatred is not prejudice if it is borne by a person who happens to hate members of one group more than he otherwise hates everyman.

201 See infra __. [Consider some further discussion of these points in Part II and then supply infra cite.]

202 In other words, the doctrine does not excuse rational discrimination. See Samuel Bagenstos, *"Rational Discrimination," Accomodation, and the Politics of (Disability) Civil Rights*, 89 Va. L. Rev. 825, 848 (2003) ("The problem of rational discrimination is a

reasons for discriminating reflect animus or hostility.[203] Thus, implicit bias research benefits antidiscrimination law by providing an empirical basis for identifying a causal connection between negative attitudes and discriminatory behaviors.

Arkes and Tetlock overestimate the value of prejudice to other judgments that we may make about behavior (such as, whether a particular behavior is discriminatory). They contend that "prejudice is a *value* as well as a *factual* judgment," but the problem with their position is that they overestimate the value prejudice as a normative component of discrimination.[204] For their restrictive definition of prejudice does not alter or undermine the legal understanding that disparate treatment resulting from feelings of guilt or avoidance that occur because of a person's protected class characteristic constitutes unlawful discrimination, whether or not such feelings are properly classified as prejudice.

Like implicit bias research itself, these criticisms of the research have migrated from social science journals to law reviews.[205] In an interdisciplinary article, Tetlock joined with Professor Gregory Mitchell to argue that implicit bias research lacks the scientific rigor to serve as a proper basis for guiding policy.[206] However, like the arguments of Arkes and Tetlock, Mitchell's and Tetlock's challenges to implicit bias research are largely normative judgments regarding the construction of implicit prejudice as a value-laden concept, and they add to those familiar challenges additional concerns about the manner in which implicit bias research may undermine the existing normative commitments of antidiscrimination law.[207] Samuel Bagenstos has rightly surmised that the arguments of Mitchell and Tetlock assume too narrow a theory of prejudice, and he reminds us that "there is no reason why it should make a difference, from an antidiscrimination law and policy perspective, whether implicit bias reflects hostility toward minorities" or some other motivation such as social awkwardness, shame or

---

central component of antidiscrimination doctrine—and it may be the most important aspect of antidiscrimination law on the ground."). *See also* Samuel Bagenstos, *Implicit Bias, "Science," and Antidiscrimination Law*, 1 HARV. L. & POL'Y REV. 481, 486-87 (2007).

[203] *See infra* Part II.B.

[204] Surely one's ignorance of just how he came to hold stereotypic beliefs about women and his inability to self-diagnose those beliefs does not let him off the hook as a misogynist when he either expresses or acts on them.

[205] *See, e.g.*, Amy L. Wax, *Discrimination as Accident*, 74 Ind. L.J. 1129 (1999) (arguing against the use of implicit bias research in antidiscrimination jurisprudence because to do so would impose overwhelming costs on employers based on behaviors beyond their reasonable control); Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 Ohio St. L.J. 1023 (2006) (arguing that implicit bias research is not "science" but value judgment and should not be adopted by law without further scrutiny).

[206] *See* Mitchell & Tetlock, *supra* note 205, at __.

[207] *See, e.g.*, *id.* at __. *See also* Bagenstos, *Implicit Bias and Antidiscrimination Law*, *supra* note 202, at 479-80 (recognizing this trajectory of Mitchell's and Tetlock's argument).

aversion.[208] But Mitchell's and Tetlock's argument is not, strictly speaking, an attempt to rationalize antidiscrimination law – it is an attempt to rationalize the concept of prejudice (including implicit prejudice) on the assumption that prejudice matters for antidiscrimination law. According to their argument, what the theory of implicit prejudice lacks is an acknowledgement of the common sense understanding that attitudes (including prejudicial attitudes) "imply an evaluative preference that, when brought to people's attention, they endorse and are even prepared to justify under appropriate conditions."[209] For Mitchell and Tetlock, a person's uncontrolled, automatic responses to racial or gender primes may indicate that they are aware of cultural stereotypes but not that they endorse them. Endorsement, they claim, is necessary to attribute a causal connection between an attitude and the kinds of controlled, deliberate – that is, intentional – behavior that is predominantly the subject of discrimination claims.

Mitchell and Tetlock are right not to confuse unconscious attitudes with unconscious behavior (as though being diagnosed with the attitude is the same as being found committing the behavior), and they are right to question whether a link between implicit biases and automatic, nonverbal behaviors (which implicit bias scholars purport to show)[210] may be transferred to intentional and even deliberate behaviors.[211] Mitchell and Tetlock aim to distinguish between "microlevel," or nonverbal, behaviors such as eye blinks and postural orientation and intentional employment actions such as subjective performance evaluations and termination decisions.[212] If this distinction is not made before lawyers adapt legal rules to fit the theory of implicit bias, the law may be asked to adopt "extraordinarily intrusive state regulation of interpersonal relations."[213] But endorsement of an attitude does not seem to speak to these questions of causality except insofar as it carries an assumption about what kinds of attitudes have sufficient motivational strength to produce action. It speaks more to the question of how we want to limit our understanding of wrongful discrimination.

On the one hand, Mitchell and Tetlock voice an empirical assumption about what makes attitudes causally strong enough to motivate intentional, deliberate actions. Psychologists will have to evaluate and debate this claim, and over time we should expect greater clarity on this issue. But, in the meantime, it would be a disproportionate response retract the scope of antidiscrimination law by requiring that intentional discrimination be demonstrated only on a showing of explicit prejudice. Particularly, as I argue, disparate treatment discrimination claims can and ought to be sustained without evidence of prejudice (in appropriate cases). On the other hand,

---

208 Bagenstos, *Implicit Bias and Antidiscrimination Law*, *supra* note 202, at 483-85.

209 Mitchell & Tetlock, *supra* note 205, at 1080.

210 [CITES]

211 Implicit bias scholars do not purport to show this kind of link, and Professor Bagenstos rightly identifies this fact as a challenge for the utility of their research in legal debates. Bagenstos, *Implicit Bias and Antidiscrimination Law*, *supra* note 202, at __.

212 Mitchell & Tetlock, *supra* note 205, at 1066-67.

213 Mitchell & Tetlock, *supra* note 205, at 1067.

Mitchell and Tetlock voice is a normative assumption about the kinds of attitudes that ought to expose a person or an institution to legal liability.

Mitchell and Tetlock are saying, at bedrock, that persons who endorse prejudicial beliefs have exposed themselves to such liability, but persons who are merely aware of the stereotypes that populate their cultural landscape should receive the benefit the doubt that they were not motivated by implicit attitudes unless one can show that they also endorsed those attitudes.[214] This view reaches beyond a critique of the science of implicit bias to a critique of the normative commitments of antidiscrimination law. I therefore agree with Bagenstos that "[t]he point of antidiscrimination law . . . is not to identify and punish prejudice as an inherent personal quality," but to remedy inequality of treatment and opportunity.[215]

B. *The normative inadequacy of prejudice as a requirement of discrimination*

The behavioral realist account interposes implicit bias as alternative to conscious prejudice (or by the terms of the realists' account, conscious intent) that, nevertheless, identifies discrimination on the basis of unjustified group-based motivations. This approach implicitly ratifies prejudice as an essential element of disparate treatment (though it does so while simultaneously enlarging the concept of prejudice to include other forms of implicit bias), and fails to entertain the possibility that intentional discrimination may occur without prejudice. This is a very different criticism from the criticisms most commonly leveled at behavioral realists and other proponents of the implicit bias literature – namely, that holding persons accountable for their implicit biases constitutes little more than "thought policing" and is motivated by political commitments to achieve particular social outcomes rather than scientific commitments that would require more rigorous proof of a connection between unconscious motivations and conscious behaviors.[216] These intuitively appealing criticisms at bottom protest that, in many instances, implicit bias should be distinguished from our traditional understanding of prejudice because a finding of implicit bias will not necessarily support the conclusion that an action is culpable. In fact, some have explicitly complained that the problem – indeed, a scientific flaw – with the inclusive marker "implicit bias" is that it exceeds the social, political and legal constructions of the concept of prejudice.[217] As

---

[214] Though Mitchell and Tetlock provide no further insight on the significance of "endorsement" for their theory of prejudice, I will consider this position in further detail in Part I.C.

[215] Bagenstos, *Implicit Bias and Antidiscrimination Law*, *supra* note 202, at 487.

[216] *See, e.g.*, Amy Wax, *Discrimination as Accident*, 75 Ind. L.J. 1129 (1999); Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 Ohio St. L.J. 1023 (2006).

[217] *See, e.g.*, Hal R. Arkes & Philip E. Tetlock, *Attributions of Implicit Prejudice, or "Would Jesse Jackson 'Fail' the Implicit Association Test?"*, 15 Psych. Inquiry 257 (2004). For example, Arkes and Tetlock distinguish between implicit bias as "shared cultural stereotypes" and "personal animus," where the latter appears to hold the ability to define

between most proponents of the implicit bias literature and this particular set of critics, the central disagreement is whether implicit bias is indeed prejudice and whether attributions of implicit bias should carry a set of consequences similar to those that have traditionally resulted from attributions of prejudice. For behavioral realists, the concept of implicit bias does update our understanding of prejudice in meaningful ways – particularly because it improves our understanding of behaviors previously understood to be caused solely by conscious prejudices. But this response misses the more fundamental point, which is that a finding of prejudice is in many circumstances not necessary to a judgment of causation or legal culpability.

For example, if an employer intentionally denies an employee meaningful information about her job performance because the employer fears that the information might be misconstrued as an expression of bias or inflates a performance evaluation because it fears that the evaluation may be tainted by bias, and the incomplete or inaccurate evaluation ultimately denies the employee an opportunity to improve her performance and to avoid other harsh consequences, has the employer not committed intentional discrimination? Neglect, appeasement, and avoidance of conflict are all behaviors that employers may commit when faced with decisions regarding the assignment, evaluation, promotion and termination of women and minorities in the workplace. If an employer commits such behaviors when dealing with women or minorities but not other workers, and the victims of these behaviors suffer adverse employment actions as a result, may the victims sustain claims of disparate treatment discrimination only if it can *also* be shown that the relevant agents of the employer harbored prejudices against the victims on the basis of their protected class characteristics? Surely it is enough that persons without their protected class characteristics were treated differently and suffered no similar adversities. By implicitly maintaining a prejudice requirement (and notwithstanding their willingness to allow that requirement to be satisfied by evidence of implicit bias), the behavioral realists have turned their back on forms of discrimination that are characterized by mere avoidance or isolation, regardless how deliberate such employer actions may be and regardless how strongly they contribute to the adversities suffered by plaintiffs.

Disparate treatment doctrine does not – and indeed should not – require proof of prejudice in order to remedy adverse employment actions that are motivated by a person's protected class status. Behavioral realists have turned to implicit bias research as though it might solve the very problem of doctrinal mismatch that it has helped to expose by expanding the scope of beliefs and attitudes that may be called "prejudiced" and thereby taken as evidence of discrimination – from those that are consciously held and deliberately pursued to those capable of motivating behavior beyond conscious awareness. But the critical question for disparate treatment is not whether the plaintiff has been the victim of either conscious or unconscious prejudice, only whether the plaintiff has been treated differently from other similarly-situated persons because of his or her protected class status. Moreover, overemphasis on prejudice may have led to the

---

the character of a person and the meaning and culpability of social behavior though the former does not. *Id*. at 258.

discouragement of a significant category of proper disparate treatment claims – claims in which the isolation, pooling or tracking of protected class members has caused them to suffer adverse treatment because of their protected class status regardless whether the employer's motivations were invidious or benevolent.[218]

The remainder of the Essay will make the case for rejecting prejudice as an absolute requirement for proof of disparate treatment. Before I undertake this argument, however, I must note that, although I deny prejudice as an absolute requirement, I do not deny its importance in particular cases. Indeed where causation is difficult to prove, evidence of prejudice may be most effective.[219]

---

[218] It is this group of discriminatory practices that has led Gary Blasi to bemoan the limitations of traditional liability-based approaches to employment discrimination and to propose the supplementation of disparate treatment litigation with a regulatory regime that would encourage employers to develop and implement "best practices" against discrimination. See supra note 90.

[219] [**To be expanded in body**.] Prejudice will be critical in cases where it is the only means that the plaintiff has to establish that the employer selected members of a particular group for adverse treatment because of their protected class status. Cases of animus-based discrimination surely qualify, but they are not the only cases in this cohort, which must also include cases where the application of legitimate evaluation or selection procedures conceals the existence of a material bias, including even from the evaluators themselves.

[May simply refer back to Part II.C. – second discussion of *Vaughn* hypothetical.] For example, an employer uses subjective evaluation procedures to rate employee performance, and an African-American employee who is terminated after receiving poor evaluations brings a disparate treatment claim. The employee does not argue that the evaluation procedures were per se discriminatory (*e.g.*, because their results or methods were not validated) or that they were a pretext for animus-based discrimination. Rather, he argues that his ratings were undeserved, and that his performance was at least as competent as several white employees who received superior evaluations. Why might this be? Perhaps the ratings were influenced by the implicit biases of the supervisors who performed the evaluations, because these supervisors undervalued his positive performance and overvalued his negative performance relative to the performance of whites. Perhaps the supervisors unconsciously harbored descriptive stereotypes of blacks that caused them to associate black employees with poor performance, and their ratings reflected that association. If the plaintiff succeeds in making his case, he will have proved that his termination was "because of" race because the performance evaluations that led to his termination were caused by the implicit prejudices of his supervisors. As an analytical matter, prejudice makes the difference; but, as a practical matter, the plaintiff's success will likely turn on his ability to show by objective standards that his performance was indistinguishable from (if not better than) that of the white employees who received superior ratings. Therefore, the key question for this kind of case is, Should the factfinder's reticence to equate this sort of under- or overvaluation of performance with prejudice present an obstacle to a finding of discrimination even though the factfinder would otherwise agree that disparities in the evaluations occurred because of race? The answer to this question must be no. Prejudice is not the required causal mechanism.

As stated above, the critical question for disparate treatment is not whether the plaintiff has been the victim of either conscious or unconscious prejudice, but whether the plaintiff has been treated differently from other similarly-situated persons because of his or her protected class characteristic. Although numerous cases discuss the plaintiff's objective in uncovering prejudice as an effort to prove discriminatory animus, neither animus nor hostility have ever been requirements of disparate treatment discrimination, and there is some Supreme Court precedent (and substantial circuit court precedent) for the conclusion that animus is specifically repudiated as a requirement under the case law.[220]

But this part of the Essay does not rest merely on a description of existing case law. Rather, I seek to make an affirmative argument for the repudiation of prejudice as a requirement. *First*, it is an obstacle – not a vehicle – to robust inclusion of the implicit bias research into the legal construction of discrimination, because negative implicit attitudes do not necessarily contain hostility toward target group members (*e.g.*, persons exhibiting implicit bias may react to members of the target group with shame, guilt, or avoidance). Moreover, a person's implicit negative attitudes may not result from a negative evaluation of others but from anxiety about how to treat others without discriminating. A requirement of prejudice may be understood as a reason to excuse disparate treatment that is the result of attitudes such as these.

Second, overemphasis on prejudice may discourage claims for which the plaintiff has no evidence of animus or hostility, even though employer practices that treat protected class members differently from nonmembers may have caused class members to suffer adverse treatment. For example, an employer may have egalitarian or pro-diversity attitudes but may also harbor some anxiety about how to avoid discrimination and achieve pro-diversity objectives. This anxiety may cause an employer to isolate protected class members by shielding them from "high profile" work as potentially high risk, even though it also contains significant opportunities for advancement or other rewards. Similarly, employers may provide protected class members with less direct feedback, or track them in a manner intended to be protective but which has the effect of restricting their advancement opportunities (e.g., by partnering them with nurturing pro-diversity mentors rather than demanding rainmakers). Should we be forbidden from categorizing such cases as examples of disparate treatment just because they are not the result of prejudice? Or because in some circumstances they may even represent failed efforts to aid or protect outgroup workers? To do so would severely limit the remedial effectiveness of employment discrimination law (particularly at a time when employers increasingly hire women and minority workers but still fail to promote such workers at rates comparable with the rates of promotion for white men).[221]

---

[220] See supra Part II [add further circuit court precedent].

[221] For many scholars, the urgency behind incorporating implicit bias research into existing doctrine concerns what researchers believe to be the new "subtle" form of discrimination, in which employers do not leave behind evidence of conscious intent to discriminate. Some scholars intend that evidence of implicit negative attitudes be used to undermine or to disprove employers' assertions that their actions are based on legitimate considerations

To a certain extent, behavioral realists have conflated modern approaches to the identification of discriminatory motivations with the conditions that define the modern workplace. Certainly, subtle discrimination motivated by implicit bias is a persistent feature of the modern American workplace. But so too are cynical and failed strategies of antidiscrimination compliance. The equal opportunity and pro-diversity initiatives of many workplaces are largely failed policies that now stand as impediments to more the development of more fair and effective institutional arrangements.[222] Antidiscrimination law should not tell persons, nominally the beneficiaries of such policies, that adverse employment actions caused by them are beyond legal scrutiny – excused by the employer's benevolent intentions.

Third, as discussed above, lay persons find attributions of prejudice more difficult than attributions of discrimination. Attributions of prejudice – particularly of racism and sexism – do not decrease in difficulty or controversy or social significance just because we may come to include unconscious attitudes as examples of prejudice.[223] What makes attributions of prejudice unappealing is the stigma perceived to attach to a person who has been identified to hold an unjustifiably biased disposition. I argue that wholesale adoption of implicit bias research as an explanation of workplace discrimination motivated by unconscious prejudices threatens to increase the social controversy of employment discrimination law because it both divorces liability from lay understandings of blameworthiness and forces factfinders to assess blameworthiness based on a more expansive definition of prejudice potentially including automatic and non-hostile unconscious attitudes. Though I agree that evidence of prejudice should continue to be probative of intentional discrimination, by maintaining a clear position that prejudice is not an absolute requirement, we may preserve for the factfinder the option of

---

and motivated by egalitarian or pro-diversity reasons. But these scholars do not directly confront the question whether employment practices should be excused from liability just because they were motivated by egalitarian values in the absence of evidence of implicit negative attitudes (*e.g.*, where the employer is unaware of unconscious hostility). I see no reason to excuse such practices if they constitute adverse action against protected class members "because of" their protected class status.

Still others may draw an analogy between such cases and circumstances in which a minority employee may bring a claim of discrimination against an employer for the stigma suffered as a result of having been the beneficiary of an affirmative action program. Setting aside the fact that Title VII explicitly authorizes the voluntary use of affirmative action programs by employers, in the cases discussed here it is not clear that the employee has benefited from receiving an inflated job evaluation or a "safe" job assignment. Moreover, employees in this hypothetical cases are not complaining of the effects of stigma, but of having been denied information or opportunities that were necessary for their continued advancement. Finally, to the extent that "stigma" caused by affirmative action programs means "prejudice" against the beneficiaries of such programs, there is nothing controversial in holding employers liable for discriminatory actions undertaken because of such prejudice.

222 Edelman.

223 Swim.

assessing the act and not the actor - to find discrimination whether or not the factfinder is able to make an attribution of prejudice.

CONCLUSION

Contemporary legal scholars propose to extend legal liability for discrimination beyond certain perceived limitations of disparate treatment doctrine by demonstrating that discriminatory behavior may be caused by implicit biases and not solely by explicit, or conscious, prejudices, and as a consequence purport to be expanding the doctrine to support claims of "subtle" discrimination. To the contrary, I conclude that the move to target discrimination caused by implicit bias does not expand the doctrine, since the doctrine already permits conduct to be judged discrimination that is motivated by unconscious attitudes or beliefs. Moreover, explicit and implicit prejudices are merely species of the same genus of unjustified beliefs or attitudes understood to cause, or motivate, discrimination. We should instead understand disparate treatment discrimination to include conduct motivated by implicit bias, and reject the notion that prejudice or conscious intent represent defining elements of disparate treatment. Rather the plaintiff's burden is fulfilled when she has shown that she received adverse treatment because of her protected class status.