

11-2010

## Liability for Unconscious Discrimination? A Thought Experiment in the Theory of Employment Discrimination Law

Patrick S. Shin

Follow this and additional works at: [https://repository.uchastings.edu/hastings\\_law\\_journal](https://repository.uchastings.edu/hastings_law_journal)



Part of the [Law Commons](#)

---

### Recommended Citation

Patrick S. Shin, *Liability for Unconscious Discrimination? A Thought Experiment in the Theory of Employment Discrimination Law*, 62 HASTINGS L.J. 67 (2010).

Available at: [https://repository.uchastings.edu/hastings\\_law\\_journal/vol62/iss1/2](https://repository.uchastings.edu/hastings_law_journal/vol62/iss1/2)

This Article is brought to you for free and open access by the Law Journals at UC Hastings Scholarship Repository. It has been accepted for inclusion in Hastings Law Journal by an authorized editor of UC Hastings Scholarship Repository.

# Liability for Unconscious Discrimination? A Thought Experiment in the Theory of Employment Discrimination Law

PATRICK S. SHIN\*

*Recent scholarship in employment discrimination law has wrestled with the problem of unconscious bias and its implications for antidiscrimination law. This Article addresses what some might regard as a naïve question: Should actions influenced by unconscious bias be regarded as discrimination under Title VII? The question might be considered naïve because any proposal for liability based on unconscious bias would surely be unripe for present implementation, and because there is no accepted method either for detecting such bias in individual cases, or for determining whether such bias actually influenced a person's decisionmaking. But these practical considerations provide no answer to the fundamental issue that underlies the question. Does unconsciously biased action fall within the legal concept of actionable discrimination? To reach that important issue, I devise a thought experiment that brackets practical problems of proof and squarely raises what I regard as the hard question for theorizing about unconscious discrimination. Should an employment action give rise to liability when that action was provably affected by the actor's unconscious bias with respect to a statutorily protected classification, even when the actor consciously acted only on legitimate, nondiscriminatory reasons? The payoff of the thought experiment is not only a clearer picture of the theoretical commitments entailed by liability based on unconscious bias, but also a keener understanding of our currently prevailing notions of actionable discrimination.*

---

\* Associate Professor of Law, Suffolk University Law School; J.D., Harvard Law School; Ph.D., Philosophy, Harvard University. I owe thanks to Larry Alexander, Frank Cooper, Mitu Gulati, Bernie Jones, Kim Krawiec, Helen Norton, Andy Perlman, Gowri Ramachandran, Al Roberts, Sam Sommers, and Bob Smith for their helpful questions and suggestions, and also to Casey Lyons and Alyssa Minsky for their excellent research assistance.

## TABLE OF CONTENTS

INTRODUCTION.....	68
II. ISOLATING THE HARD QUESTION .....	70
A. TWO ASSUMPTIONS.....	72
1. <i>The First Assumption</i> .....	72
2. <i>The Second Assumption</i> .....	74
B. A HYPOTHETICAL TEST CASE: “WORK EXPERIENCE” .....	75
III. LEGAL BACKGROUND .....	77
A. CASE LAW .....	77
B. THE LITERAL STATUTORY ARGUMENT FOR LIABILITY .....	80
C. SUMMARY.....	83
IV. THE NORMATIVE QUESTION .....	83
A. SURFACE ARGUMENTS.....	83
B. THE DEEPER PROBLEM.....	85
V. CAUSAL AND JUSTIFICATORY CONCEPTIONS OF DISCRIMINATION.....	86
A. WHY THE LAW RESISTS LIABILITY FOR UNCONSCIOUS DISCRIMINATION.....	88
1. <i>Current Predominance of the Justificatory Conception</i> .....	89
2. <i>Liability for Unconscious Discrimination and the             Causal Conception</i> .....	91
3. <i>Precepts of Responsibility</i> .....	93
C. IMPLICATIONS OF A COMMITMENT TO THE CAUSAL CONCEPTION .....	95
1. <i>The Charge of Discrimination as Diagnosis</i> .....	96
2. <i>The Problem of Conceptual Instability: “Work             Experience II”</i> .....	97
3. <i>Biting the Bullets</i> .....	99
CONCLUSION .....	100

## INTRODUCTION

A steadily mounting body of social science research provides strong evidence that our intentional actions are often influenced by psychological factors that are not present to our introspective awareness—for example, “implicit” or unconscious biases.<sup>1</sup> One direct implication of this research is that ascertaining a person’s conscious

---

1. For a review of the literature, see Katherine Bartlett, *Making Good on Good Intentions: The Critical Role of Motivation in Reducing Implicit Workplace Discrimination*, 95 VA. L. REV. 1893, 1904–20 (2009), and Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CALIF. L. REV. 945 (2006).

motives for an action may not always provide a complete explanation of why she did it. The most comprehensive explanation of an action might require positing the influence of a psychological factor that played no part in the actor's conscious deliberations. Thus, an actor's decision to take a particular adverse action against one individual rather than another might be best understood as having been affected by the actor's awareness of the race of those individuals, even if the actor could sincerely disavow any conscious consideration of race as a reason for the decision.

The possibility of this sort of scenario—cases in which an objectionable bias affects an action without entering the actor's own awareness—presents a worrisome impediment to the achievement of justice in the workplace. There have been numerous deeply insightful articles describing the problem of implicit bias and discussing its implications for antidiscrimination law.<sup>2</sup> My purpose in this Article is to focus directly on what might be called a naïve question of liability: should implicit bias be a basis of disparate treatment liability under Title VII?<sup>3</sup> The question might fairly be regarded as naïve insofar as any proposal for such liability would surely be unripe for present implementation in light of serious issues pertaining to problems of proof in individual cases, not to mention intramural disputes among experts about the proper practical inferences that can be drawn from studies of implicit bias. My interest, however, is more fundamental. I want to understand whether and how the notion of unconsciously biased action should figure into our current legal conception of actionable discrimination. This is a thoroughly normative, not a scientific or epistemological, question.<sup>4</sup>

The question of this Article, then, is whether employment actions causally affected by implicit bias should *ipso facto* be regarded as

---

2. See generally Samuel R. Bagenstos, *The Structural Turn and the Limits of Antidiscrimination Law*, 94 CALIF. L. REV. 1, 5–7 (2006); Bartlett, *supra* note 1; Tristin K. Green, *Discrimination in Workplace Dynamics: Toward a Structural Account of Disparate Treatment Theory*, 38 HARV. C.R.-C.L. L. REV. 91 (2003); Christine Jolls & Cass R. Sunstein, *The Law of Implicit Bias*, 94 CALIF. L. REV. 969 (2006); Jerry Kang & Mahzarin R. Banaji, *Fair Measures: A Behavioral Realist Revision of "Affirmative Action,"* 94 CALIF. L. REV. 1063 (2006); Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161 (1995); Ann C. McGinley, *Viva la Evolucion!: Recognizing Unconscious Motive in Title VII*, 9 CORNELL J.L. & PUB. POL'Y 415 (2000); David Benjamin Oppenheimer, *Negligent Discrimination*, 141 U. PA. L. REV. 899 (1993); Susan Sturm, *Second Generation Employment Discrimination: A Structural Approach*, 101 COLUM. L. REV. 458 (2001); Amy L. Wax, *Discrimination as Accident*, 74 IND. L.J. 1129 (1999).

3. Title VII of the Civil Rights Act of 1964 § 703(a), 42 U.S.C. § 2000e-2(a) (2006).

4. See Samuel R. Bagenstos, *Implicit Bias, "Science," and Antidiscrimination Law*, 1 HARV. L. & POL'Y REV. 477, 491 (2007) ("Science does not defeat the implicit bias law-reform program, but science does not establish the case for that program, either. That program depends on a normative judgment that discrimination is not about fault but about a social problem—a normative judgment that is deeply contested among judges and policymakers.").

actionably discriminatory, even when the actor genuinely and reasonably believed that the action was justified by nondiscriminatory considerations. Part of what is really at stake in the question of liability for unconsciously biased action is an implicit judgment about the relative priority that should be given to the sufficiency of an agent's non-pretextual rationale for the action, versus socio-psychological explanations that deemphasize the importance of that deliberative perspective and focus instead on the action's causal etiology. I argue that the most significant conceptual shift entailed by liability for truly "unconscious discrimination" is not the shift from a fault-based to a consequence-based standard of liability (the shift on which others have focused),<sup>5</sup> but from an agent-relative, justificatory conception to what might be described as a diagnostic, causally-oriented understanding of what constitutes discriminatory action. I contend that this shift would represent a significant departure from, rather than an incremental expansion of, current conceptions of discrimination, and I discuss concerns about whether a causal conception of discrimination might diminish the moral significance of the charge of discrimination and whether it may create a deep instability for any distinction between individual discriminatory action and "societal" discrimination.

## II. ISOLATING THE HARD QUESTION

The question whether employment discrimination laws should allow for liability on a theory of implicit bias is, in a sense, purely academic. It is not clear that practicable, scientifically accepted methods for proving that a particular individual harbors implicit bias, or for establishing that any particular action in non-laboratory conditions was affected by such bias, currently exist.<sup>6</sup> This is not to deny the convincing evidence of implicit bias (although it is not entirely uncontroversial to claim that measures of implicit bias, like the Harvard Implicit Association Test (IAT),<sup>7</sup> are actually predictive of any real-world action).<sup>8</sup> But it is one

---

5. *See id.*

6. *See* Amy L. Wax, *The Discriminating Mind: Define It, Prove It*, 40 CONN. L. REV. 979, 981–86 (2008).

7. *See* Bagenstos, *supra* note 2, at 6 (describing the test).

8. For analysis in support of the predictive value of the IAT, see Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity*, 97 J. PERSONALITY & SOC. PSYCHOL. 17 (2009). For a study that raises some critical questions, see Andrew Karpinski & James L. Hilton, *Attitudes and the Implicit Association Test*, 81 J. PERSONALITY & SOC. PSYCHOL. 774 (2001). For an overtly skeptical perspective, see Hal R. Arkes & Philip E. Tetlock, *Attributions of Implicit Prejudice, or "Would Jesse Jackson 'Fail' the Implicit Association Test?"*, 15 PSYCHOL. INQUIRY 257, 264, 270 (2004) (objecting to reaction time data on implicit bias on the grounds that it may reflect cultural stereotypes instead of prejudice, negative responses may be due to emotions unrelated to prejudice, and that sometimes data showing prejudice can also show rational behavior), and Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 OHIO ST. L.J. 1023, 1023 (2006) (arguing that research on implicit bias needs to

thing to accept the research that supports the hypothesis that people act from biases of which they are not consciously aware; it is arguably quite another to adduce proof sufficient to establish that unconscious bias affected a specific act of decisionmaking in the complex setting of the workplace, even in the face of the actor's genuine disavowals.<sup>9</sup>

Thus, it would not be unfair to take the position that the question whether the employment discrimination laws should allow for liability based on proof of implicit bias is unripe for practical and meaningful discussion at present.<sup>10</sup> This difficulty would seem to counsel strongly in favor of reform proposals that address the problem of implicit bias without trying to make it a basis for liability under Title VII or other antidiscrimination statutes.<sup>11</sup> For example, in their well-known article, Jerry Kang and Mahzarin Banaji argue that implicit bias research provides a basis for the use of affirmative action as a way to introduce "debiasing" agents into the workplace,<sup>12</sup> but they are careful to disclaim any argument for outright liability.<sup>13</sup>

More generally, the current impracticability of a liability scheme premised on implicit bias might be thought to obviate any inquiry about whether such liability could be justified. But I think that is the wrong conclusion to draw. In fact, the concern about the unripeness of the question of liability implies a distinction between the practical and the theoretical question. The point of the concern is that the practical difficulties provide reason to put off addressing the theoretical issue of liability. The concern implicitly acknowledges the existence of a further question, masked by problems of scientific validity, evidentiary proof, and doctrinal implementation.<sup>14</sup> What I want to do is to bracket those

---

include a more "rigorous investigation of . . . error rates," an analysis of the predictive utility of measuring implicit bias in the workplace, and a discussion of social consequences of changing requirements for proving bias). See generally Bruce Bower, *The Bias Finders*, 169 SCI. NEWS 250 (2006) (reporting disagreement among psychologists about what exactly can be inferred from IAT findings).

9. See Linda Hamilton Krieger & Susan T. Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 CALIF. L. REV. 997, 1034-36 (2006).

10. Cf. Melissa Hart, *Subjective Decisionmaking and Unconscious Discrimination*, 56 ALA. L. REV. 741, 774-77 (2005) (discussing practical difficulties); Mitchell & Tetlock, *supra* note 8, at 1067-70 (asserting the claimed link between unconscious discrimination and discriminatory behavior have "little legal significance"); Franita Tolson, *The Boundaries of Litigating Unconscious Discrimination: Firm-Based Remedies in Response to a Hostile Judiciary*, 33 DEL. J. CORP. L. 347, 421 (2008) (observing practical challenges to solutions proposed by scholars).

11. For examples of such proposals, see Ivan E. Bodensteiner, *The Implications of Psychological Research Related to Unconscious Discrimination and Implicit Bias in Proving Intentional Discrimination*, 73 MO. L. REV. 83, 108-28 (2008) (suggesting procedural litigation reforms to help account for implicit bias), Green, *supra* note 2, at 144-56 (calling for "structural" workplace reform as a response to the problem of implicit bias), and Sturm, *supra* note 2, at 479-522, 553-61 (same).

12. Kang & Banaji, *supra* note 2, at 1108-15.

13. *Id.* at 1077 ("We are not arguing that implicit bias-induced discrimination should produce the same legal liability as explicit animus-driven discrimination under current . . . federal antidiscrimination statutes.").

14. Cf. Bagenstos, *supra* note 4, at 492 (arguing that some forms of skepticism about the legal

practical difficulties in order to reach the normative issue of whether the employment discrimination laws should be understood to encompass liability for actions tainted by unconscious bias.

#### A. TWO ASSUMPTIONS

To that end, I am going to assume the truth of some controversial hypotheses. In doing so, I do not mean to dismiss disagreement about them as unimportant or illusory. The purpose of these assumptions is to isolate and expose the central normative questions about how the law should regard actions affected by implicit bias. My intent is to disentangle my arguments from debates that can only play out in the social scientific realm. The payoff of this discussion, I hope, will be not only a clearer picture of the theoretical commitments entailed by liability based on unconscious bias, but a keener understanding of our currently prevailing notion of actionable discrimination.

##### 1. *The First Assumption*

The first controversial hypothesis that I am going to assume to be true is that unconscious or implicit bias is *real*. To avoid any misunderstanding about exactly what this assumption is supposed to mean,<sup>15</sup> I shall take some time to explain. At a minimum, the assumption implies that people can have psychological states or dispositions of which they are not conscious,<sup>16</sup> including unconscious dispositions affecting actual behavior.<sup>17</sup> I take this to entail that it is possible for a true and accurate description of a person's psychological dispositions to include preferences or other proclivities of which the person is unaware and is unable to make himself aware.<sup>18</sup> It also means that it is possible for a

---

implications of implicit bias research are really normative arguments about the scope of actionable discrimination, and therefore calling for "a renewed attention to antidiscrimination *theory*").

15. For an expansive and insightful examination of the meaning of "unconscious" and the question whether the unconscious "exists," see MICHAEL S. MOORE, *LAW AND PSYCHIATRY: RETHINKING THE RELATIONSHIP* 126-47, 249-80 (1984).

16. I do not venture any claim about the metaphysics of psychological or mental states, and I daresay my discussion does not depend on any particular thesis about the nature of the relation between mental states and physical brain states.

17. Here, again, I wish to avoid any entanglement in philosophical debates about the metaphysics of mental causation. I hope it will suffice to stipulate that I am taking a broadly functionalist approach to the nature of mental states—an approach under which we can understand mental or psychological states in terms of their functional roles in thought and behavior, independent of any underlying thesis about how such states might be realized in or reduced to physical brain states. See MOORE, *supra* note 15, at 35.

18. I add "unable to make himself aware" to distinguish mental states of which a person might not presently be conscious, but of which the person could become aware if she made an effort to direct her attention to them. This sort of temporarily latent mental state, for example, a suppressed state of hunger, is not unconscious in the sense that makes implicit bias problematic. *Cf. id.* at 130 (using the Freudian term "preconscious" to refer to this simple sense of "unconscious"). In the sense of unconscious that is relevant to the problem of implicit bias, to say that we had certain unconscious mental states means that "we are not able to recall them at all, even if we do direct considerable

person to act from a particular disposition without being aware, and without being able to become aware, that she has acted from that disposition.<sup>19</sup>

The assumption that implicit bias is “real” implies, more specifically, that a person can have an unconscious psychological disposition that comprises or results in a bias toward certain types of individuals as compared to others. That is, a person can have an unconscious psychological state that comprises or results in a disposition to act more favorably toward individuals with certain characteristics than individuals without those characteristics—which is to say that those characteristics can influence that person’s actions such that she would act differently in the absence of that influence. Finally, and crucially, the assumption implies that a person may act from this kind of unconscious disposition even when she could genuinely disavow consideration of the disposing characteristic.<sup>20</sup> In other words, the person’s action might be influenced by that characteristic without the actor being aware of that influence.<sup>21</sup>

I will use the term “unconscious discrimination” to refer to this sort of hypothesized case—in other words, where an agent acts from an unconscious bias inaccessible to her own awareness—and in which the disposing characteristic at issue is one that is a statutorily forbidden basis for employment action (race, color, religion, sex, national origin, age, disability).<sup>22</sup> The assumption that implicit bias is real means, therefore, that unconscious discrimination is possible.

To make this more concrete, consider the example of unconscious discrimination on account of race. What does it mean to say that race-based implicit bias is “real” and that unconscious discrimination because of race is possible? What I take this to mean is that some people have unconscious biases that dispose them to act more favorably to members of certain races than to members of others, and, furthermore, that these unconscious racial biases sometimes actually affect how people act.<sup>23</sup>

attention to the question of what they were.” *Id.* at 131.

19. Michael Moore suggests, in a similar vein, that one thing that we might mean when we say that a person has a particular “unconscious mental state” is “that the holder of the mental state does not have the capacity to recognize the state that he is in; he cannot describe it even if he attempts to direct his attention to it.” *Id.* at 129.

20. For a discussion of empirical support for this possibility, see David L. Faigman et al., *A Matter of Fit: The Law of Discrimination and the Science of Implicit Bias*, 59 *HASTINGS L.J.* 1389, 1404–06 (2008) (“[S]elf-generated explanations of one’s own thought processes are often no more accurate than that of outside observers who have little knowledge of the mental content of another person.”).

21. For a concise overview of the social science research providing evidence of the prevalence of this sort of “dissociation” between self-professed attitudes and behavior, see Gary Blasi, *Default Discrimination: Law, Science, and Unintended Discrimination in the New Workplace*, in 3 *NYU SELECTED ESSAYS ON LABOR AND EMPLOYMENT LAW: BEHAVIORAL ANALYSES OF WORKPLACE DISCRIMINATION* 3, 6–12 (Mitu Gulati & Michael J. Yelnosky eds., 2007).

22. Title VII of the Civil Rights Act of 1964 § 703(a), 42 U.S.C. § 2000e-2(a) (2006).

23. Michael Selmi has noted that the term “unconscious discrimination” is vague and could

What does it mean for an unconscious racial bias to actually affect how a person acts? The most straightforward case would be an agent taking an action with respect to an individual of a particular race that the agent would not have taken if only the individual was of a different race—thus constituting bias. Yet, an honest report by the agent of the considerations on which she believed she acted would not include the individual's race (this is the warrant for regarding the bias as unconscious).

The assumption that unconscious bias is real entails a claim about the possible motivation of action generally. This claim is that an agent's actions are sometimes influenced by biases of which the agent is unaware. If the claim is true, there seems little reason to doubt that unconscious bias affects actions by decisionmakers in a wide variety of contexts,<sup>24</sup> including the workplace. From the perspective of employment discrimination laws, the potential concern is evident. Employment actions, no less than any other kind of action, might be influenced by decision makers' unconscious biases relating to race, color, religion, sex, national origin, age, or disability—the considerations excluded from permissible consideration under Title VII,<sup>25</sup> the Age Discrimination in Employment Act (ADEA),<sup>26</sup> and Title I of the Americans with Disabilities Act (ADA).<sup>27</sup> The question that interests me is whether actions that suffer from this sort of bias—acts of unconscious discrimination—fit within the concept of actionable discrimination embedded in these legal frameworks.

## 2. *The Second Assumption*

The second big assumption of this Article is that unconscious discrimination is *provable*—in other words, that the influence of implicit bias on an agent's action is something that can, in principle, be proved in individual cases.<sup>28</sup> This assumption is presently probably closer to science fiction than a plausible supposition, given what I understand to be the

---

potentially be used to refer to any kind of “subtle” discrimination not manifested in the form of overt racism. See Michael Selmi, *Response to Professor Wax: Discrimination as Accident: Old Whine, New Bottle*, 74 IND. L.J. 1233, 1236 (1999). I am not using the term in this loose way, but rather to refer specifically to differential treatment influenced by the psychological operation of implicit bias of which the actor is unaware.

24. See, e.g., Alafair S. Burke, *Improving Prosecutorial Decision Making: Some Lessons of Cognitive Science*, 47 WM. & MARY L. REV. 1587, 1603–05 (2006); Antony Page & Michael J. Pitts, *Poll Workers, Election Administration, and the Problem of Implicit Bias*, 15 MICH. J. RACE & L. 1, 21–33 (2009); Jeffrey J. Pokorak, *Probing the Capital Prosecutor's Perspective: Race of the Discretionary Actors*, 83 CORNELL L. REV. 1811, 1818–19 (1998); Samuel R. Sommers, *On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations*, 90 J. PERSONALITY & SOC. PSYCHOL. 597, 606–09 (2006).

25. 42 U.S.C. § 2000e-2(a).

26. 29 U.S.C. § 623(a) (2006).

27. 42 U.S.C. § 12112(a) (2006 & Supp. II).

28. To put it another way, I assume that the influencing effect of implicit bias on action is a “datable event.”

current state of the relevant sciences.<sup>29</sup> Be that as it may, I am going to suppose that it is, in principle, possible to detect or rule out the influence of unconscious bias with respect to any given action by an individual agent, the agent's genuine disavowals notwithstanding.<sup>30</sup> I will imagine, more specifically, that it is in principle possible to determine whether, in a given case, the agent's unconscious bias was such that it caused the agent to act differently than she would have in the absence of the bias. In short, I am assuming that the truth of a claim that a particular action constituted unconscious discrimination is something that can be determined, on a preponderance of the evidence, through empirical methods of proof.<sup>31</sup>

This second assumption, much more than the first, will obviously require a temporary suspension of practical skepticism. As I said at the outset, the point of the assumption is not to deny or minimize the possibility that it might in fact be false. It is, rather, to enable us to drill down to the basic normative issue of liability that might otherwise be clouded by worries about practical implementation.

#### B. A HYPOTHETICAL TEST CASE: "WORK EXPERIENCE"

I want to construct a hypothetical set of facts that puts my two assumptions to work and provides us with a concrete case that squarely presents what I have been characterizing as the basic normative issue of liability for unconscious discrimination. I will refer to this scenario as the "Work Experience" case.

An employer wants to hire someone for a managerial position. The employer has to decide between two candidates. Both are qualified and

29. See Blasi, *supra* note 21, at 10–11; Wax, *supra* note 6, at 985–86.

30. If one likes, one might imagine that a battery of psychological tests and brain-imaging techniques could be developed for this purpose. Cf. Blasi, *supra* note 21, at 10 (describing studies linking perception of racial difference with certain patterns of brain activation).

31. On this assumption, unconscious bias is a discrete, detectable psychological phenomenon, as opposed to a conceptual construct that merely stands in for the inexplicability of an action under more intentional descriptions. The difference between these types of views is given some elaboration by Alasdair MacIntyre in his discussion of Freud's theory of the unconscious. See A.C. MACINTYRE, *THE UNCONSCIOUS: A CONCEPTUAL ANALYSIS* 50–79 (1958). As MacIntyre puts it, "Either the unconscious is an inaccessible realm of inaccessible entities existing in its own right or it is a theoretical and unobservable entity introduced to explain and relate a number of otherwise inexplicable phenomena." *Id.* at 71. In his discussion of MacIntyre's view of Freud, Thomas D. D'Andrea nicely summarizes the latter type of view of the unconscious in this way:

[T]he unconscious . . . represents simply an abductive inference to a better explanation of the causes of certain forms of overt human behaviour. In positing an unconscious mind (or unconscious processes at least) . . . Freud's inference . . . conforms to a standard pattern of scientific explanation: one which seeks to link observable to observable via an unobservable process.

THOMAS D. D'ANDREA, *TRADITION, RATIONALITY, AND VIRTUE: THE THOUGHT OF ALASDAIR MACINTYRE* 167 (2006). For further philosophical discussion about the possible meanings of "unconscious," see MOORE, *supra* note 15, at 126–42.

have roughly comparable credentials, except that one applicant, who is black, performed slightly better in the job interview, while the other applicant, who is white, has slightly more work experience. In her own private deliberations about which applicant would be the best candidate for the job, the employer decides that work experience is the most important factor. The employer therefore chooses the white applicant. Although the employer is aware of each applicant's race, she honestly believes that her choice was based on the white applicant's superior work experience, and she can honestly deny any conscious consideration of the candidates' race in her decisionmaking. However, despite the employer's honest belief, her decision was in fact influenced by unconscious bias in favor of whites and against blacks. More specifically, if it had been the black applicant who had more work experience and the white applicant who had done better in the interview, the employer would still have selected the white applicant, but she would have done so on the basis of her superior performance in the interview;<sup>32</sup> and she would have genuinely believed that this was the basis of her choice.

The truth of the matter in the "Work Experience" case is, in short, that the employer had an unconscious disposition to disfavor the black applicant. She acted from that disposition, all the while believing that she was acting on a reason that had nothing to do with the race of the applicants, when, in fact, her professed reason for action was really just a secondary, epiphenomenal rationalization that effectively shielded her own bias from introspective discovery. Yet, the employer's professed reliance on the chosen applicant's credentials is not a conscious cover or *pretext* for discrimination, because she genuinely believed it was her reason for acting. But, by hypothesis, it is the subconscious influence of the employer's awareness of the race of the applicants that actually explains her action.

In conjunction with these facts, let us imagine also that testing methods or other modes of proof exist that allow us to determine that despite the employer's good faith disavowal of bias, and despite her subjective belief that she acted on a perfectly legitimate consideration in choosing the white candidate, her awareness of the race of the black applicant did in fact (more likely than not) influence her decision to select the white candidate instead, such that the employer would have selected the white candidate over the black one even if their application dossiers had been switched.<sup>33</sup>

---

32. This hypothetical scenario is loosely inspired by actual research studies. See Michael I. Norton et al., *Mixed Motives and Racial Bias: The Impact of Legitimate and Illegitimate Criteria on Decision Making*, 12 PSYCHOL. PUB. POL'Y & L. 36, 41-42 (2006); Eric Luis Uhlmann & Geoffrey L. Cohen, *Constructed Criteria: Redefining Merit to Justify Discrimination*, 16 PSYCHOL. SCI. 474, 475, 477-78 (2005).

33. Larry Alexander constructs a similar hypothetical case of unconscious race-based preference in the context of a broad and comprehensive discussion of the moral wrongness of discrimination generally. See Larry Alexander, *What Makes Wrongful Discrimination Wrong? Biases, Preferences,*

The foregoing hypothesized facts establish a scenario in which an employer's selection of a job candidate is provably affected by unconscious bias. My two assumptions allow us to discuss this scenario without getting mired in internecine disputes about implicit bias or practical proof problems. Important matters, to be sure, but setting them aside allows us to see clearly the hard question for theorizing about unconscious discrimination: Should an employment action—like our hypothetical employer's selection of the white job candidate—give rise to employment discrimination liability when the action was provably affected by the actor's unconscious bias with respect to a statutorily protected classification, even when the actor consciously acted only on legitimate, nondiscriminatory reasons? This is the hard question because it lays bare the question of how the law should respond to provable unconscious discrimination. Given the imagined facts in the "Work Experience" case, should the employer be subject to liability under the employment discrimination laws?

### III. LEGAL BACKGROUND

I approach the question primarily as a normative, theoretical one. But before moving to that question, This Part briefly discusses how a claim of unconscious discrimination might fare under current case law and whether such a claim would fit the relevant statutory text.

#### A. CASE LAW

An argument for liability on the hypothesized facts of the "Work Experience" case probably would not succeed under currently accepted judicial frameworks of analysis. First, it is unlikely that our hypothetical black job candidate would prevail on a disparate treatment claim under Title VII,<sup>34</sup> which requires a showing of discriminatory intent or motive.<sup>35</sup>

---

*Stereotypes, and Proxies*, 141 U. PA. L. REV. 149, 180 (1992). Alexander's imagined case involves a sports fan whose unconscious bias causes his allegiances to shift toward whichever hometown team happens to be the most predominantly white, even though he does not believe he is biased and in fact "rejects all biased judgments at the conscious level." *See id.*

34. The basic statutory cause of action for discrimination on a theory of disparate treatment is provided in § 703(a) of the Civil Rights Act of 1964, 42 U.S.C. § 2000e-2(a)(1) (2006). The framework for analyzing the sufficiency of a plaintiff's claim under the statute is set forth in the Supreme Court's decisions in *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 802–05 (1973), and its progeny. *See, e.g.,* *Reeves v. Sanderson Plumbing Prods., Inc.*, 530 U.S. 133, 142–43 (2000); *St. Mary's Honor Ctr. v. Hicks*, 509 U.S. 502, 506–12 (1993); *Tex. Dep't of Cmty. Affairs v. Burdine*, 450 U.S. 248, 252–56 (1981).

35. Justice O'Connor stated this point unequivocally in *Reeves*, although the significance of unconscious bias was not an issue in that case: "The ultimate question in every employment discrimination case involving a claim of disparate treatment is whether the plaintiff was the victim of intentional discrimination." 530 U.S. at 153; *see also* Blasi, *supra* note 21, at 17 ("The fundamental schema of anti-discrimination laws is borrowed from the law of intentional torts: plaintiff victims of discrimination are permitted to sue defendant employers for damages if they can establish disparate

More specifically, the rejected candidate might be able to make out a prima facie case,<sup>36</sup> but the employer's assertion that the hiring of the white applicant was based on consideration of the applicant's work experience would constitute a nondiscriminatory rationale that would rebut the inference of discrimination generated by the prima facie case.<sup>37</sup> Ultimately, the plaintiff's success would depend on proof that the defendant's stated reason was merely a pretext for discrimination.<sup>38</sup> On the assumed facts of the "Work Experience" case, however, that proof would not be possible, as we are stipulating that the employer can honestly assert that she consciously based her decision on a comparative evaluation of work experience.<sup>39</sup> In short, our hypothetical employer is not subject to liability on any conventional theory of intentional discrimination.<sup>40</sup> This should hardly be surprising<sup>41</sup> and, indeed, is part of

---

treatment 'because of' the employee's race, sex, or other protected category.'). For arguments that disparate treatment liability does not necessarily require proof of conscious discriminatory intent, see *infra* note 41.

36. See *Int'l Bhd. of Teamsters v. United States*, 431 U.S. 324, 358 n.44 (1977) (explaining that in general, the requirement of the prima facie case merely requires plaintiff to demonstrate that the challenged employment action did not result from plaintiff's lack of qualifications or the absence of a job vacancy); see also *Swierkiewicz v. Sorema, N.A.*, 534 U.S. 506, 512 (2002) (stating that the prima facie case should be understood flexibly).

37. See *Reeves*, 530 U.S. at 148.

38. See *Hicks*, 509 U.S. at 511.

39. See, e.g., *Udo v. Tomes*, 54 F.3d 9, 13-14 (1st Cir. 1995) (applying the *McDonnell Douglas* framework to reject plaintiff's claim on grounds that plaintiff failed to establish that improper motive resulted in the layoff and therefore, that plaintiff's evidence did not support an inference of discrimination); see also Tristin K. Green, *A Structural Approach as Antidiscrimination Mandate: Locating Employer Wrong*, 60 VAND. L. REV. 849, 877-80 (2007); Jolls & Sunstein, *supra* note 2, at 980. See generally *Millbrook v. IBP, Inc.*, 280 F.3d 1169, 1175 (7th Cir. 2002); Krieger & Fiske, *supra* note 9, at 1034-38 (describing and criticizing the "honest belief" rule).

40. Nor is there any basis, on the hypothesized facts, for disparate treatment liability under a "pattern or practice" theory, under which an inference of discriminatory intent could be drawn from statistically significant racial patterns in an employer's hiring or promotion practices. See *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 309-12 (1977); *Teamsters*, 431 U.S. at 339-40. It is unclear whether the imagined facts of "Work Experience" would be sufficient to make it a "mixed motives" case subject to the "motivating factor" liability standard articulated in § 703(m) of Title VII of the Civil Rights Act, 42 U.S.C. § 2000e-2(m) (2006). However, I would think that a court that understood the case law to require proof of conscious consideration of a protected characteristic as a basis for disparate treatment liability would probably interpret the "motivating factor" language of § 703(m) to refer to conscious motives. This is simply a claim about how I would expect most courts to read the statute if the issue of unconscious discrimination were squarely presented. As I argue below in Part III.B., the literal statutory text of Title VII arguably does not *require* such a reading and, in fact, seems open to the possibility of liability based on unconscious bias.

41. Some commentators have argued that the cases do not clearly support the claim that disparate treatment claims depend on proof of conscious bias, animus, or consideration of a protected characteristic. In an article published in 2000, Ann McGinley proposed a reading of the case law under which unconscious bias could constitute pretext where the employer's decisionmaker is "mistaken" about the relevant facts, such as the qualifications of the affected employees or job candidates. See McGinley, *supra* note 2, at 453-56. Katharine Bartlett has also argued that the cases do not uniformly support the claim that disparate treatment claims depend on proof of discriminatory intent in the sense of conscious bias or animus. See Bartlett, *supra* note 1, at 1922-24.

the point of our imagined scenario.<sup>42</sup>

Second, it is similarly unlikely that the rejected black candidate in the “Work Experience” case would have a conventional disparate impact claim. Although Title VII disparate impact claims, unlike disparate treatment claims, do not require proof of discriminatory intent,<sup>43</sup> they do require that the plaintiff identify “a particular employment practice”<sup>44</sup> that creates a significant disparate impact.<sup>45</sup> In our example, neither of those elements is present. It is (for all we know) a singular case of adverse treatment not tied to any particular employment practice as such,<sup>46</sup> and there is no observed disparate impact apart from the effect of the employer’s decision on the two managerial candidates under discussion. It might be possible that the aggregate consequences of the employer’s discretionary decisionmaking procedure over time would reflect a pattern that could support a disparate impact claim,<sup>47</sup> but that possibility depends on facts that go beyond my hypothetical. The question is not whether we could add further facts that would support a claim of discrimination, but whether the hypothesized singular decision

42. There is some authority, at least in the First Circuit, that unconscious reliance on discriminatory stereotypes may be sufficient to establish the discriminatory intent required for a disparate treatment claim. See *Swallow v. Fetzer Vineyards*, 46 F. App’x 636, 644 (1st Cir. 2002); *Thomas v. Eastman Kodak Co.*, 138 F.3d 38, 58 (1st Cir. 1999); *Small v. Mass. Inst. of Tech.*, 584 F. Supp. 2d 284, 294 (D. Mass. 2008) (stating that circumstantial evidence could be used to show pretext). It seems to me an interesting question whether the notion of an unconscious reliance on a racial stereotype is different from implicit bias generally. Cf. Selmi, *supra* note 23, at 1241 (suggesting that not all stereotypes may influence behavior). I would note that a discriminatory stereotype is typically perceived of as an illegitimate belief about attributes of members of the stereotyped class. Thus, the concept of unconsciously acting on a stereotype may centrally involve the idea of an *unconscious belief*. Depending on our understanding of belief, this sort of unconscious bias seems arguably distinguishable from, or at least a special case of, unconscious bias as I have conceptualized it, namely, as a functional state or disposition, as opposed to a cognitive state with propositional content. See *supra* notes 15–21 and accompanying text. In any event, the unconscious stereotype cases are not particularly apposite to the present discussion, because the facts of “Work Experience” are not meant to suggest that the employer’s action there is based on any reliance, unconscious or otherwise, on any stereotype as such. For further discussion of this point, see *infra* text accompanying notes 85–90.

43. As Justice Kennedy explained in his opinion for the Court in *Ricci v. DeStefano*, “Title VII prohibits both intentional discrimination (known as ‘disparate treatment’) as well as, in some cases, practices that are not intended to discriminate but in fact have a disproportionately adverse effect on minorities (known as ‘disparate impact’).” 129 S. Ct. 2658, 2672 (2009); see also *Teamsters*, 431 U.S. at 335–36 n.15 (explaining the difference between disparate treatment and disparate impact claims).

44. 42 U.S.C. § 2000e-2(k)(1)(A)(i).

45. See *Ricci*, 129 S. Ct. at 2678 (characterizing the prima facie case of disparate impact liability to require “essentially, a threshold showing of a significant statistical disparity”); *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 994–95 (1988); *Connecticut v. Teal*, 457 U.S. 440, 446 (1982).

46. The statute excuses the plaintiff from showing that a particular employment practice caused the disparate impact at issue in certain circumstances not relevant here. See 42 U.S.C. § 2000e-2(k)(1)(B)(i) (“[I]f the complaining party can demonstrate to the court that the elements of a respondent’s decision-making process are not capable of separation for analysis, the decisionmaking process may be analyzed as one employment practice.”).

47. See *Watson*, 487 U.S. at 989–96; cf. *Regents of Univ. of Cal. v. Bakke*, 438 U.S. 265, 309 n.44 (1978).

to select the white candidate instead of the black candidate in the “Work Experience” case itself constitutes actionable discrimination.

#### B. THE LITERAL STATUTORY ARGUMENT FOR LIABILITY

Although my central concern is with the normative question of liability for unconscious discrimination, it is worth pausing, to make an observation about the statutory language that forms the basis for the disparate treatment cause of action under Title VII. I stated above that an argument for liability on the facts of the “Work Experience” hypothetical would be implausible under current case law. Such an argument does not seem to be foreclosed, however, by the literal language of Title VII.<sup>48</sup> If one were to approach the statute afresh rather than through the lens of Supreme Court precedent, one might very well read it to permit the imposition of liability for unconscious bias.<sup>49</sup>

The central liability provision of Title VII says that it is unlawful “to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual . . . because of such individual’s race, color, religion, sex, or national origin.”<sup>50</sup> Liability is thus predicated on “discriminat[ion] . . . because of” a protected characteristic.<sup>51</sup> This language implies that in order for liability to be imposed, there must be a particular sort of relation between the adverse employment action in question (the “discrimination”) and a protected characteristic of the aggrieved individual, such as race—to wit, a relation captured by the phrase “because of.”<sup>52</sup>

The term “discriminate” is undefined in Title VII. The statutory language does not expressly say that an action must involve conscious consideration of a factor before it can constitute discrimination because

---

48. See Bartlett, *supra* note 1, at 1922; Wax, *supra* note 6, at 982–84; cf. McGinley, *supra* note 2, at 447; Wax, *supra* note 2, at 1146.

49. My argument shows that liability for unconscious discrimination is not precluded by the relevant statutory text. I make no claim about what Congress might have intended with regard to unconscious discrimination.

50. 42 U.S.C. § 2000e-2(a).

51. *Id.* The relevant language of the ADEA is similar. See Age Discrimination in Employment Act, 29 U.S.C. § 623(a) (2006) (“It shall be unlawful for an employer . . . to . . . discriminate against any individual . . . because of such individual’s age.”). Interestingly, the Americans with Disabilities Act of 1990 originally contained the same “because of” language in its general liability provision, but as part of the ADA Amendments Act of 2008, that “because of” language was replaced with the phrase “on the basis of.” See ADA Amendments Act of 2008, Pub. L. No. 110-325, § 5(a), 122 Stat. 3553 (codified as amended at 42 U.S.C. § 12112 (2006 & Supp. II 2008)).

52. Similarly, § 703(b) of Title VII, which is sometimes understood to provide textual support for disparate impact liability, makes it unlawful for an employer “to limit, segregate, or classify his employees or applicants for employment in any way which would deprive or tend to deprive any individual of employment opportunities or otherwise adversely affect his status as an employee, *because of* such individual’s race, color, religion, sex, or national origin.” 42 U.S.C. § 2000e-2(b) (emphasis added).

of that factor. Although at times courts seem to suggest that discrimination as such is inherently intentional,<sup>53</sup> those claims tend to be made in the context of distinguishing actionable discrimination from differential treatment that is merely correlated with a proscribed consideration,<sup>54</sup> not in the context of distinguishing the legal significance of unconscious rather than conscious bias. Indeed, since it is uncontroversially true that Title VII's conception of discrimination encompasses liability for disparate impact, which does not depend upon any showing of intentional consideration of any protected classification, it seems impossible to read the term "discriminate" to preclude actions influenced by unconscious bias.

The semantics of the "because of" construction seem similarly open to the possibility of liability for unconscious discrimination. In ordinary usage, when we say that a person acted "because of" some factor, we are saying that the factor explains why the person so acted. But there is nothing in the "because of" construction that necessarily implies that the explanation must be in terms of the person's conscious intention: the construction "A did x because of y" does not necessarily imply that y was A's conscious rationale for doing x.<sup>55</sup> It only implies, in ordinary usage, that y is a consideration that helps explain A's doing of x, a consideration that tells us something about why A did x.<sup>56</sup> No semantic distortion is required to assert that the employer in our hypothetical "Work Experience" case chose the white candidate "because of" race. The literal terms of § 703(a) of Title VII, which prohibit "discriminat[ion] . . . because of"<sup>57</sup> a protected characteristic, need not be read to apply only in cases where an employer's action is explained by the conscious consideration of race or another protected characteristic.<sup>58</sup>

53. *See, e.g.*, *United States v. North Carolina*, 914 F. Supp. 1257, 1265 (E.D.N.C. 1996) ("The concept of 'unintentional discrimination' is logically impossible.").

54. *See Hazen Paper Co. v. Biggins*, 507 U.S. 604, 611–12 (1993).

55. Consider the sentence, "Smith flew to Florida because of y." To be sure, y might be some factor, like the warm weather, that Smith consciously considered. But it would make just as much sense for y to be something wholly outside Smith's conscious deliberation—for example, "Smith flew to Florida because he boarded the wrong plane;" or even, "Smith flew to Florida because he was under the spell of an evil demon." Furthermore, there is nothing about the "because of" construction that implies that the proposition "A did x because of y" must hold true for one and only one specification of y. It might be true, for example, that Smith went to Florida because he wanted to be in warm weather; but it might at the same time be true that Smith went to Florida because it evokes pleasant memories of childhood vacations with his parents.

56. In a recent case interpreting the "because of" construction as it appears in the ADEA, Justice Thomas asserted that "because of" entails a but-for causal relation. *Gross v. FBL Fin. Servs., Inc.*, 129 S. Ct. 2343, 2350–51 (2009). It is true that specifying a causal relation between x and y is one way of explaining x in terms of y. But as I argue below, it is not the only way.

57. 42 U.S.C. § 2000e-2(a).

58. There are some reported decisions in which courts have suggested that an employer cannot be held liable for discriminating against an employee "because of" a protected characteristic unless the employer was actually aware that the employee possessed that characteristic. *See, e.g.*, *Geraci v.*

The language of the alternative proof standard articulated in § 703(m) of Title VII,<sup>59</sup> generally thought to apply to “mixed motive” cases,<sup>60</sup> likewise does not seem to preclude liability for unconscious discrimination. Under § 703(m), liability is established upon proof that an impermissible classification was “a motivating factor for any employment practice, even though other factors also motivated the practice.”<sup>61</sup> This language by itself does not preclude liability for unconscious discrimination, so long as the “motivation” of an action is understood to include unconscious as well as conscious influences that bear on an employer’s decisionmaking.<sup>62</sup> If we understand motivation broadly to refer to an action’s actual psychological impetus, it does not seem implausible to think of unconscious bias as something that can “motivate” action.<sup>63</sup> Thus, when an employer’s action is affected by unconscious bias with respect to a prohibited consideration, it is semantically plausible to claim the illegitimate consideration is “a motivating factor” which would be sufficient to establish liability under § 703(m).

In the “Work Experience” case, it is stipulated that the race of the two managerial candidates explains the employer’s action, even though

---

Moody-Tottrup, Int’l, Inc., 82 F.3d 578, 581 (3d Cir. 1996); Hedberg v. Ind. Bell Tel. Co., 47 F.3d 928, 932–33 (7th Cir. 1995). The reason for rejecting liability in such cases, however, is that the employee’s protected characteristic cannot possibly make a difference to an employer’s decision and, therefore, cannot help explain it if that characteristic is not even known to the employer. The requirement of awareness of the protected characteristic does not speak to whether the employer must be *consciously motivated* by that awareness.

59. 42 U.S.C. § 2000e-2(m).

60. There is presently considerable uncertainty about the circumstances in which the “motivating factor” standard of § 703(m) governs the sufficiency of a Title VII plaintiff’s proof of discrimination. The Supreme Court in *Desert Palace, Inc. v. Costa* rejected the imposition of a heightened evidentiary requirement as a precondition to the application of § 703(m). 539 U.S. 90, 101–02 (2003). At the same time, the Court assumed without deciding that § 703(m) is properly applied only to a certain subset of disparate treatment cases (“mixed motive” cases), while other cases would presumably be governed by the proof framework originally laid out in *McDonnell Douglas*. *See id.* at 94 n.1. The *Desert Palace* Court declined, however, to provide positive guidance as to how the subset of cases governed by § 703(m) should be delineated. The Court’s silence has given rise to a divergence of approaches among the circuit courts. *Compare* *Ginger v. District of Columbia*, 527 F.3d 1340, 1345–46 (D.C. Cir. 2008) (refusing to apply § 703(m) where plaintiff failed to argue a mixed motives theory of liability), *with* *Dominguez-Curry v. Nev. Transp. Dep’t*, 424 F.3d 1027, 1037–42 (9th Cir. 2005) (applying the “motivating factor” standard of § 703(m) to a case in which the plaintiff had relied on *McDonnell Douglas* to establish her prima facie case). *See generally* Michael J. Zimmer, *A Chain of Inferences Proving Discrimination*, 79 U. COLO. L. REV. 1243 (2008) [hereinafter Zimmer, *A Chain of Inferences*] (analyzing the relationship between *McDonnell Douglas* and the § 703(m) standard of proof); Michael J. Zimmer, *The New Discrimination Law: Price Waterhouse is Dead, Whither McDonnell Douglas?*, 53 EMORY L.J. 1887 (2004) [hereinafter Zimmer, *The New Discrimination*] (discussing the logical implications of *Desert Palace* for structures of proof in disparate treatment litigation).

61. 42 U.S.C. § 2000e-2(m).

62. *See* Faigman et al., *supra* note 20, at 1397 (noting Congress’s silence on whether implicit bias can constitute “a motivating factor”).

63. *See id.* at 1395–97.

the employer did not consciously consider that factor. On this assumption, it seems perfectly plausible to say that the employer chose the white individual over the black one “because of such individual’s race,”<sup>64</sup> and also that the candidates’ race was “a motivating factor for”<sup>65</sup> the employer’s decision. The literal language of Title VII does not foreclose holding the employer liable, even though she never acted on any conscious consideration of race.

### C. SUMMARY

I have suggested that under existing case law, it would be difficult to argue for employer liability on the facts of the “Work Experience” case, even though a literal reading of the relevant text of Title VII does not foreclose such an argument. Perhaps the more relevant observation, though, is that courts have not had occasion to consider squarely the precise issue that the hypothetical case presents. In the non-hypothetical world, practical issues of proof and scientific validity make the question of liability based on unconscious bias alone unripe for litigation and judicial consideration.<sup>66</sup> Thus, even if some case law could be understood as allowing for the possibility of employer liability in our “Work Experience” hypothetical, there is no authority that would unequivocally support that result. In what follows, I turn my attention to the question of whether the law ought to allow for such liability. Should our employment discrimination laws permit the imposition of employer liability based on proof of unconscious discrimination? What reasons are there in favor of such liability? And what reasons might there be for the law to remain inhospitable to such claims?

## IV. THE NORMATIVE QUESTION

In this Part, I summarize various familiar arguments in favor of and against adapting Title VII to account for unconscious discrimination. I then explore why those “surface” arguments fail to provide a satisfactory answer to the deeper normative question whether unconsciously biased action should be regarded as legally actionable discrimination.

### A. SURFACE ARGUMENTS

The arguments in favor of interpreting or reforming Title VII to account for unconscious discrimination are familiar.<sup>67</sup> These arguments

---

64. 42 U.S.C. § 2000e-2(a).

65. *Id.* § 2000e-2(m).

66. See Selmi, *supra* note 23, at 1243 (“[T]here is a veritable absence of litigation over the unconscious nature of discrimination, an issue that is rarely raised in reported cases.”).

67. For a concise survey of some of the reforms that have been proposed by scholars, see Bartlett, *supra* note 1, at 1926–30.

generally share the premise that unconscious bias not only is real in the sense described above, but that it is pervasive.<sup>68</sup> If that is so, then we should expect it to be a significant source of workplace inequality and unequal treatment, a source of inequality no less important than intentional discrimination or the sort of practices that produce outcomes cognizable as disparate impact. The basic argument, given this expectation, is simple. If unconscious bias really does pervade decisionmaking in the workplace, then it will be impossible to achieve the most basic goals of employment discrimination law—to create genuine equality of opportunity and to eliminate patterns of substantive inequality that track protected categories<sup>69</sup>—unless we adapt the law to be able to respond to such bias.<sup>70</sup>

Whether or not the employment discrimination laws can effectively reduce the prevalence of bias,<sup>71</sup> one thing they can do is to make local corrections for resultant inequalities by allowing for the imposition of liability where it can be proved that a forbidden bias—conscious or unconscious—explains the adverse decision at issue. For example, by invalidating the employer’s decision in the “Work Experience” case, the law could correct an economic distribution (the allocation of the job to the white candidate) that is traceable to an adverse racial attitude. This is precisely the sort of distributive inequality that the employment discrimination statutes are meant to address.<sup>72</sup> A commitment to achieving this basic goal of employment discrimination law<sup>73</sup> thus creates a reason in favor of extending its reach to include employment actions influenced by unconscious bias.<sup>74</sup>

---

68. See generally Greenwald & Krieger, *supra* note 1; Kang & Banaji, *supra* note 2, at 1072 (analyzing implicit bias and its pervasiveness).

69. See *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 800–01 (1973); *Griggs v. Duke Power Co.*, 401 U.S. 424, 430–31 (1971).

70. See Krieger, *supra* note 2, at 1241; see also Eva Paterson et al., *The Id, the Ego, and Equal Protection in the 21st Century: Building Upon Charles Lawrence’s Vision to Mount a Contemporary Challenge to the Intent Doctrine*, 40 CONN. L. REV. 1175, 1195 (2008).

71. See Christine Jolls, *Antidiscrimination Law’s Effects on Implicit Bias*, in 3 NYU SELECTED ESSAYS ON LABOR AND EMPLOYMENT LAW: BEHAVIORAL ANALYSES OF WORKPLACE DISCRIMINATION, *supra* note 21, at 69, 73.

72. Cf. *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 990–91 (1988) (arguing that “subconscious stereotypes and prejudices” that infect “undisciplined” subjective decision making are a “lingering form of the problem that Title VII was enacted to combat” and should be subject to Title VII liability, insofar as such decisionmaking “has precisely the same effects as a system pervaded by impermissible intentional discrimination”).

73. See *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 421 (1975) (noting that a central purpose of Title VII is “making persons whole for injuries” resulting from discrimination).

74. The argument for liability need not exclude the possibility that an employer could voluntarily undertake measures to reduce the effects of bias. See, e.g., Tristin K. Green & Alexandra Kalev, *Discrimination-Reducing Measures at the Relational Level*, 59 HASTINGS L.J. 1435, 1438–54 (2008) (discussing steps to reduce unconscious discrimination that may reduce or insulate underlying levels of implicit bias).

On the other side, there are a variety of arguments against liability for unconscious discrimination. Some of them seem predicated primarily on skepticism about the actuality of unconscious discrimination<sup>75</sup> or on the practical and epistemological difficulties of establishing that a particular action was in fact influenced by unconscious bias.<sup>76</sup>

Other arguments against liability for unconscious discrimination raise empirical concerns about efficacy. For example, one might contend that holding employers liable for unconscious discrimination will not significantly improve workplace equality stemming from implicit bias because even if such liability might provide corrective action in a tiny number of litigated cases, it would not effectively reduce the actual incidences of implicit bias in the employment context.<sup>77</sup> In addition, liability for unconscious discrimination could actually have adverse consequences for minorities in the workplace. Such liability might increase the predicted costs of hiring employees who are members of a statutorily protected group.<sup>78</sup> Further, perceptions that such liability would be unfairly punitive could create a counterproductive backlash of resistance and negative attitudes toward the law, which could impede the internalization of antidiscrimination norms by workplace actors.<sup>79</sup>

#### B. THE DEEPER PROBLEM

All of the foregoing arguments, though, miss a deeper point of contention, a potential source of reluctance to embrace the possibility of liability for unconscious discrimination that does not simply boil down to skepticism about the efficacy of such a regime in achieving desired goals and does not depend on disproving that such liability would in fact reduce workplace inequality. In other words, I think that liability for unconscious discrimination would remain controversial *even if* there were conclusive evidence that it would in fact reduce workplace inequality overall. The question I want to explore is: why?

From a skeptical perspective, all of the arguments in favor of liability for unconscious discrimination are question-begging in a crucial way. They presuppose that something objectionable happens whenever an employer action was causally influenced by a protected characteristic of the adversely affected individual, regardless of whether the employer genuinely believed she had adequate reasons for the action. In other

---

75. For references to some of the critical literature, see Bartlett, *supra* note 1, at 1896 n.4.

76. See Wax, *supra* note 6, at 981–86. These worries are largely mooted by my working assumption that unconscious discrimination is real.

77. See Jolls & Sunstein, *supra* note 2, at 986–87.

78. Cf. Christine Jolls, Commentary, *Antidiscrimination and Accommodation*, 115 HARV. L. REV. 642, 686–87 (2001) (arguing that antidiscrimination laws impose greater costs to employers because they must employ certain employees against customer and coworker attitudes).

79. See Bartlett, *supra* note 1, at 1936–41.

words, the arguments assume that if a protected characteristic is part of the causal predicate that explains what happened, then a corrective response is in order. This is surely a plausible position, but it assumes an answer to the hard question at hand. The hard question is precisely whether, as a matter of discrimination law, we *should* believe something objectionable occurred when, though an employer genuinely sought to act on permissible—even justifiable—reasons, the employer’s actions were influenced by an employee’s protected characteristic.

Consider again the “Work Experience” hypothetical. Suppose that we were completely convinced that, from a systemic perspective, a regime of liability for unconscious discrimination was necessary to achieve meaningful reductions of race-based inequality in employment. Would we then have to conclude, as a matter of logic, that the employer’s decision in the “Work Experience” case should be treated as *actionable discrimination*? I do not believe so, and the reason is that policy arguments of the sort canvassed above do not reach the deepest point of controversy. That question is whether we should regard the employer’s action as legally invalid simply because it was precipitated by implicit bias, *even though* the reasons on which the employer consciously sought to act were in fact legally adequate to justify the decision that she made. Claiming that the decision should be legally invalid because it had a race-related cause simply begs the critical question of what, fundamentally, the legal charge of discrimination entails. Should the objection entail only that the action in question was causally influenced by a protected characteristic? Or should it entail that the action’s justification was invalid or defective in virtue of incorporating an illegitimate consideration? Should the legal charge of discrimination be understood fundamentally as a demand for justification of action, or as a call for a causal inquiry into an historical event?

#### V. CAUSAL AND JUSTIFICATORY CONCEPTIONS OF DISCRIMINATION

The question, in short, is whether employment discrimination law should embrace a conception of discrimination that hinges on examination of the causal history of an employment event or, in contrast, on the putative justificatory rationale of the employer’s action. The first view—call it a “causal” conception of discrimination—would incline toward a finding of liability in the “Work Experience” case because it is clear that the rejected candidate’s race, a protected characteristic, is necessary to a complete explanation of what actually brought about the observed result.<sup>80</sup> The other view—call it a “justificatory” conception—

---

80. See Krieger, *supra* note 2, at 1242 (“The critical inquiry would be whether the applicant or employee’s group status ‘made a difference’ in the employer’s action, not whether the decisionmaker intended that it make a difference.”).

would probably find no actionable discrimination in the “Work Experience” hypothetical, because no forbidden consideration played any role in the justificatory rationale that informed the employer’s own understanding of her decision about which candidate to hire.

Doubts about liability for unconscious discrimination on the facts of the “Work Experience” case, even with all of the accompanying stipulations, come from the perspective of a justificatory conception of discrimination. I argue in this Part that the justificatory conception is engrained in our current law, and that the causal conception would represent a radical shift in our understanding of what discrimination is, not just an incremental expansion of the scope of the employment discrimination laws.

What I am calling the deep distinction between the causal and justificatory conceptions of discrimination does not simply reduce to a question about whether discrimination should be restricted to intentional acts of differential treatment on the basis of a forbidden consideration, or whether the law should also impose liability for *unintentional* acts of differential treatment. The unease we may feel about liability for unconscious discrimination is not explained by some general objection to liability in the absence of intent to discriminate. After all, the law already provides a cause of action—the disparate impact claim—that allows the imposition of liability for discrimination even in the *absence* of any intention to treat individuals differently on account of a protected characteristic. Because our employment discrimination law already countenances liability in the absence of discriminatory intent, the deep objection to liability for unconscious discrimination cannot be based on a normative difficulty relating to that consideration. The problem with liability for unconscious discrimination must be something else.

If the law already accepts the possibility of liability in the absence of discriminatory intent, why should it have a problem with liability for unconscious discrimination? My claim is that the difficulty with liability for unconscious discrimination is that such liability depends on a causal conception of discrimination that displaces the significance of a justificatory one. Returning to the facts of the “Work Experience” case may help to illustrate the difference. We have been focusing on the stipulated fact that the employer’s decision to hire the white job candidate was precipitated by an unconscious psychological disposition to disfavor blacks (or to favor whites). But there is another important assumed fact: namely, that the employer’s subjective motivation in acting—her belief that the white applicant was the better candidate because of her work experience—is nondiscriminatory and, what is more, seems to reflect an objectively adequate reason for selecting that candidate. Superiority of work experience is a legitimate consideration that can provide sufficient reason for preferring one job candidate over

another—in other words, a permissible reason for the sort of decision in question.

We can gain an important insight by asking what principle is implied by the claim that the employer's selection of the white applicant in the "Work Experience" case should be regarded as actionable discrimination. I argue that it implies that the objective justifiability of an action from the perspective of the actor's genuine understanding of the relevant considerations matters less than an agent-neutral inquiry into the factors that had a causal influence on the ultimate outcome. The causal inquiry trumps and displaces the justificatory one for purposes of determining whether an action constitutes discrimination. In other words, if our hypothetical employer is liable, that means the objective adequacy of her own genuine reasons-based explanation of her action must yield to the causal psychological history that culminated in that act. The legal status of the act depends on its psychological etiology, not on the actor's ability to satisfy a demand for justification.

This causal conception of actionable discrimination allows for the imposition of liability even when the employer acted on considerations that provide legitimate, adequate reasons for the adverse differential treatment in question. I contend that this conception implicates a fundamental change in the meaning of discrimination. It changes the charge of discrimination from an action-guiding demand of equal respect to a factual assertion about the psychology of the actor and transforms the relevant inquiry from the quasi-moral business of calling upon the employer to justify her action to the quasi-scientific business of identifying the psychological causal antecedents of that action. In a regime of liability for unconscious discrimination, the statutory proscription against discrimination shifts from an action-guiding principle that constrains the considerations that can justify differential adverse treatment to a kind of mandatory tax on unconscious contributions to the problem of workplace inequality.

#### A. WHY THE LAW RESISTS LIABILITY FOR UNCONSCIOUS DISCRIMINATION

Is the shift to a causal conception of discrimination really a radical change from the law's current understanding? I believe it is. Consider again the currently prevailing models of actionable discrimination. In a disparate treatment action, an employer can respond to the charge of discrimination by articulating a nondiscriminatory reason for the adverse action at issue. Unless the aggrieved employee can prove that this articulated reason was pretextual, the employer's stated justification for the adverse action will defeat the charge.<sup>81</sup> Thus, the charge of discrimination implies that either there were no nondiscriminatory

---

81. See *St. Mary's Honor Ctr. v. Hicks*, 509 U.S. 502, 509–11 (1993).

reasons supporting the employer's adverse action,<sup>82</sup> or that the employer acted on considerations that are statutorily proscribed. Conversely, if there was a nondiscriminatory reason for the challenged action, and if that reason genuinely constituted the employer's rationale for that action, then there is no actionable discrimination on a disparate treatment theory.<sup>83</sup> The focus of the inquiry is on whether the employer's reasons—the employer's rationale for acting—passes muster under a principle of nondiscrimination.<sup>84</sup> The statutory proscription serves as an action-guiding constraint on the considerations that can count as valid reasons for differential adverse treatment.

### *I. Current Predominance of the Justificatory Conception*

Contemporary theories of disparate treatment based on social stereotyping also fit within the justificatory conception of discrimination, even though they might at first blush seem to incorporate causal elements. In a stereotyping case, the plaintiff seeks to establish the intent element of the disparate treatment claim by an allegation that the employer acted on an illegitimate generalization about members of a protected class.<sup>85</sup> Treating a person differently because of such a stereotype is clearly a form of actionable discrimination.<sup>86</sup> Insofar as prejudicial stereotypes that affect our judgments about others may be largely inchoate and even subconscious, it might seem that liability based on stereotypes is akin to liability based on unconscious discrimination<sup>87</sup>—therefore, disproving my claim that the causal conception of discrimination is foreign to our current understanding. Liability based on prejudicial stereotypes, however—whether conscious or unconscious<sup>88</sup>—cannot be reduced to a purely causal inquiry into whether a psychological attitude influenced the actor's cognition. A stereotype is

---

82. In which case, the employer could not meet its burden of production in the standard *McDonnell Douglas* burden-shifting framework. See *Tex. Dep't of Cmty. Affairs v. Burdine*, 450 U.S. 248, 255 (1981).

83. In a mixed-motives case, the plaintiff only need prove that one of the statutorily forbidden considerations was "a motivating factor" in the employer's action, and the employer can be held liable even if other considerations also motivated that action. Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e-2(m) (2006). This is typically understood to require some proof that the employer took a statutorily forbidden consideration into account, even though it also relied on other factors in taking the adverse action at issue. See *Desert Palace, Inc. v. Costa*, 539 U.S. 90, 98–99 (2003). Thus, even in a mixed-motives case, there is no actionable discrimination unless a discriminatory reason constitutes at least part of the employer's own true understanding of the basis for its action.

84. See, e.g., *Burdine*, 450 U.S. at 254–56.

85. See *Price Waterhouse v. Hopkins*, 490 U.S. 228, 241–42 (1989).

86. See *id.* at 239.

87. See Zimmer, *A Chain of Inferences*, *supra* note 60, at 1279 ("[E]vidence of stereotypical thinking supports an ultimate inference of intent to discriminate precisely because it is an unconscious expression of bias.").

88. See *Thomas v. Eastman Kodak Co.*, 138 F.3d 38, 58 (1st Cir. 1999).

arguably a form of propositional belief, a schematic construct<sup>89</sup> that provides putative reasons for action or judgment.<sup>90</sup> To act on a stereotype is to act, consciously or unconsciously, on the basis of a belief about members of a particular class—e.g., “all *X*’s have property *F*” or “all good *Y*’s exhibit behavior *G*.” The legal invalidation of actions based on such stereotypes entails a judgment that beliefs of that sort are not to be regarded as legitimate reasons for taking an adverse employment action. Holding an employer liable for acting on a prejudicial stereotype constitutes a rejection of a putative *rationale*—namely, the stereotyping belief—for the action, a refutation of the action’s true justification. The same cannot be said for holding the employer liable for the sort of unconscious discrimination illustrated in the “Work Experience” case, because the employer’s action, by hypothesis, does not entail any sort of racial stereotyping beliefs, conscious or otherwise. There, the predicate for liability is the causal influence of a racial attitude (a negative disposition toward blacks) in the employer’s formation of her rationale for action—in other words, that work experience is more important than interview performance, not the objectionable nature of the rationale itself as a principle of action.

Finally, even the disparate impact cause of action evinces a justificatory conception of actionable discrimination, not a causal one. While it is of course true that disparate impact liability does not require proof of intent to discriminate,<sup>91</sup> the disparate impact cause of action must still ultimately be conceived as an evaluation of the employer’s rationale in carrying out whatever practices created the observed disparities in question. This is so because the employer can avoid liability by proving that the particular employment practice that created the observed disparate outcomes was “job related . . . and consistent with business necessity.”<sup>92</sup> In other words, although the employer rather than the plaintiff bears the burden of proving business necessity, the important point is that the employer can avoid liability, even in a disparate impact case, by showing that the challenged employment practice was justified by adequate reasons.<sup>93</sup> Although the plaintiff can establish a prima facie case of disparate impact by showing that the employer’s practices *caused* a certain pattern of inequality,<sup>94</sup> liability

---

89. See Krieger, *supra* note 2, at 1195 (“Stereotypes are correlational constructs.”).

90. See Patrick S. Shin, *Vive la Différence? A Critical Analysis of the Justification of Sex-Dependent Workplace Restrictions on Dress and Grooming*, 14 DUKE J. GENDER L. & POL’Y 491, 499–501 (2007) (discussing different types of behavior leading to stereotypes).

91. *Ricci v. DeStefano*, 129 S. Ct. 2658, 2672 (2009).

92. Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2006).

93. See *Ricci*, 129 S. Ct. at 2678–81 (explaining that liability for disparate impact ultimately depends on the employer’s inability to prove that its practices were justified by business necessity, or a less discriminatory alternative to those practices was available).

94. See *id.* at 2677–78.

depends upon a justificatory inquiry into whether the employer's practices were necessary: Were those practices justified by good business reasons, and were they justified even in light of other available alternatives?<sup>95</sup> If the practices were justified, then the employer prevails, even though the employer's practices might continue to cause racially disparate consequences.<sup>96</sup> Disparate impact liability therefore depends ultimately on a test of the adequacy of the employer's putative rationale. In this context, no less than in the case of disparate treatment, the prohibition of discrimination can be seen as an action-guiding demand for justification, a requirement that employment practices that adversely affect members of protected groups be backed by a legally sufficient reason.

## 2. *Liability for Unconscious Discrimination and the Causal Conception*

Consider, in contrast, the conception of actionable discrimination implied by the possibility of liability in the "Work Experience" hypothetical. In that case, there seems to be an acceptable justification for the decision to hire the white applicant (a preference for work experience), and this justification constitutes the employer's ostensive rationale, her genuine understanding of the basis of her action. We cannot say that the employer lacked adequate justification for hiring the white applicant, nor can we say that the decision was unjustified under the employer's rationale. A view under which this employer's decision is regarded as actionably discriminatory effectively rejects the idea that we should give controlling significance to the question of the adequacy of the action's rationale. What is controlling is that the employer's reasoning, whether or not it was sufficient to justify the action, was in fact influenced by her awareness of the race of the two applicants.<sup>97</sup>

Moreover, if the employer is liable because of the operation of unconscious bias, then, in an important way, the proscription against discrimination ceases to be an action-guiding constraint on employment action.<sup>98</sup> In the "Work Experience" case, by hypothesis, the employer did not treat any illicit consideration as a reason in favor of her decision, and the bias that influenced her act does not represent any attitude or

---

95. *See id.* at 2678–81.

96. *See id.*

97. Recall that we are assuming that the employer would have reasoned to the opposite conclusion but for the rejected applicant's being black.

98. As Linda Krieger has described the argument, "the normative utility of a rule prohibiting discrimination depends entirely on decisionmaker self-awareness. . . . Absent decisionmaker self-awareness, the nondiscrimination principle—if framed solely as a prohibitory injunction 'not to discriminate'—loses its normative mooring." Krieger, *supra* note 2, at 1186. Professor Krieger argues, of course, that discrimination should not be defined in terms of a violation of that sort of prohibitory injunction. *See id.* at 1239–40.

judgment that the employer himself would endorse. If liability is imposed in cases like the “Work Experience” hypothetical, then the statutory proscription becomes more akin to what might be regarded as a tax, a mandatory cost levied against the operation of the employer’s unconscious bias, regardless of the employer’s ability to justify her action under nondiscriminatory principles.

The conceptual shift implicated by liability for unconscious discrimination, then, is a shift from a prescriptive, action-guiding regime that constrains the considerations that can provide legitimate reasons for adverse action to a primarily diagnostic regime that focuses on ferreting out potential psychological causes of persisting workplace inequality. The prohibition against discrimination changes from a demand that employers be prepared to justify their actions on grounds that do not involve a protected characteristic to a statement putting employers on notice that actions that adversely affect members of protected groups will give rise to liability upon proof that the actions were causally influenced by such a characteristic. This is a move from a quasi-moral, justification-based conception of discrimination to a quasi-scientific, psychological-causal conception.

It is true that, in a way, our current conception of discrimination is already partly psychological and depends, as legal responsibility does more generally, on proof of causation. Disparate treatment liability depends, for example, on an examination of what a decisionmaker actually considered and believed, and that sort of inquiry is in some sense a matter of identifying the considerations that were causally relevant in the actor’s actual psychology.<sup>99</sup> So, I am certainly not denying that causation is relevant to the law’s current conception of discrimination, nor am I trying to make any sort of broad claim about the notions of causation that figure into legal doctrines governing responsibility more generally.<sup>100</sup> My claim is that the argument in favor of liability for unconscious discrimination departs radically from our current conception of discrimination, because it requires cleaving the causation inquiry entirely from any evaluation of the adequacy of the agent’s subjective rationale for the action in dispute.<sup>101</sup>

---

99. See Zimmer, *A Chain of Inferences*, *supra* note 60, at 1244 (discussing the requirement that the plaintiff prove a “link” between the defendant’s discriminatory motive and the adverse employment action at issue).

100. For a careful and exhaustive inquiry into the concepts of causation that are implicit and explicit in doctrines of responsibility in criminal law and torts, see generally MICHAEL S. MOORE, *CAUSATION AND RESPONSIBILITY: AN ESSAY IN LAW, MORALS, AND METAPHYSICS* (2009).

101. Sheila Foster has made the point, albeit in the service of a much different thesis than I am propounding here, that “the causal inquiry in antidiscrimination cases [under current law] is evaluative, not explanatory.” Sheila R. Foster, *Causation in Antidiscrimination Law: Beyond Intent Versus Impact*, 41 Hous. L. Rev. 1469, 1517 (2005) (emphasis omitted) (cautioning that determinations about the causes of a decision or set of consequences may *themselves* be subject to the influence of

It is also true that the current legal framework governing disparate treatment claims requires a causal inquiry relating to the actor's psychology in the limited sense that liability depends upon figuring out what beliefs, perceptions, and judgments actually constituted the rationale on which the employer relied in acting as she did.<sup>102</sup> But on the model of discrimination that would impose liability for unconscious bias, the inquiry bypasses the actor's putative rationale and seeks an explanation of the act in terms of *any* aspect of the actor's psychology—not just the beliefs and attitudes that the actor himself would endorse—that could be regarded as a causal influence. What is radical here is the implicit premise that whenever an unconscious bias can be identified as a causal influence, liability is justified, even if the actor's own rationale for acting provides adequate, nondiscriminatory reasons for the action in question. On the causal conception, when inquiries into the actor's genuine rationale and the scientifically determinable psychological causes of her action yield different answers to the question whether a protected characteristic made a difference to the actor's decision, then it is the scientific psychological cause, not our putative reason-giving, that controls whether the action should be regarded as objectionable discrimination.

### 3. *Precepts of Responsibility*

What explains the felt reluctance to embrace liability for unconscious discrimination is, I think, the philosophical tension between (a) conceiving of an action as a product of scientifically determinable, subterranean psychological causes; and (b) conceiving of that action as the sort of thing for which an agent should be held responsible in any robust sense.<sup>103</sup> In the context of practical debates about whether to hold a person responsible for an act, to explain a given act as the determined consequence of antecedent historical causes normally has the force of preempting claims of moral responsibility for the act.<sup>104</sup> This tendency should be familiar from basic philosophical debates about criminal

---

implicit biases and stereotyping beliefs).

102. *Tex. Dep't of Cmty. Affairs v. Burdine*, 450 U.S. 248, 255–56 (1981).

103. This general tension is one of the central themes of the philosophy of action or action theory. Seminal contemporary works that identify many of the questions that have occupied theorists in this field include G.E.M. ANSCOMBE, *INTENTION* (Harvard Univ. Press 2d ed. 2000), DONALD DAVIDSON, *ESSAYS ON ACTIONS AND EVENTS* (1980), and Peter Strawson, *Freedom and Resentment*, in *FREE WILL* 59–80 (Gary Watson ed., 1982). For an accessible discussion of how some of these questions apply in the context of criminal law, see R.A. DUFF, *INTENTION, AGENCY & CRIMINAL LIABILITY: PHILOSOPHY OF ACTION AND THE CRIMINAL LAW* (1990).

104. Note that I am making a claim about the force of deterministic causal explanations in the context of *practical deliberation* about whether to hold a person responsible for an act. I am not suggesting, and in fact would deny, that causal explanations in general are incompatible with attributions of responsibility. For philosophical perspectives that inform my own views on this latter question, see T.M. SCANLON, *WHAT WE OWE TO EACH OTHER* 248–94 (1998); R. JAY WALLACE, *RESPONSIBILITY AND THE MORAL SENTIMENTS* 51–83 (1994).

responsibility.<sup>105</sup> We might be reminded of Chief Justice Weintraub's concurring opinion in *State v. Sikora*.<sup>106</sup> The issue for the court there had to do with the legal relevance of expert psychiatric testimony opining that a criminal defendant's act of murder was "motivated by the predetermined influence of his unconscious"<sup>107</sup> and therefore lacked the level of intentionality required for a conviction of first degree murder.<sup>108</sup> More interesting for present purposes than the court's holding on this question<sup>109</sup> is the philosophical question taken up by Chief Justice Weintraub in his concurrence. His concurrence is notable because of how strikingly, yet to opposite effect, his discussion relates to the question of liability for unconscious discrimination:

The psychiatric view [of the defendant's criminal act] seems quite scientific. It rests upon the elementary concept of cause and effect. The individual is deemed the product of many causes. As a matter of historical fact, he was not the author of any of the formative forces, nor of his capacity or lack of capacity to deal with them. . . .

. . . .

. . . [The psychiatrist] traces a man's every deed to some cause truly beyond the actor's own making, and says that although the man was aware of his action, he was unaware of assembled forces in his unconscious which decided his course. Thus the conscious is a puppet, and the unconscious the puppeteer.<sup>110</sup>

Characterizing the psychiatrist's view as likening the criminal defendant to an "automaton whose unconscious directs . . . antisocial deeds,"<sup>111</sup> Justice Weintraub argues that this view is "simply irreconcilable" with the very possibility of criminal responsibility:

To grant a role in our existing [legal] structure to the theme that the conscious is just the innocent puppet of a nonculpable unconscious is to make a mishmash of the criminal law, permitting—indeed requiring—each trier of the facts to choose between the automaton thesis and the law's existing concept of criminal accountability. It would be absurd to decide criminal blameworthiness upon a psychiatric thesis which can find no basis for personal blame. So long as we adhere to criminal blameworthiness, *mens rea* must be sought and decided at the level of conscious behavior.<sup>112</sup>

Justice Weintraub sees the causal model of action (the picture of the conscious actor as the puppet of unconscious motives) as antithetical to

105. See generally DUFF, *supra* note 103, at 101–02.

106. 210 A.2d 193, 204–07 (N.J. 1965).

107. *Id.* at 201 (majority opinion).

108. See *id.* at 202.

109. The court concluded that the psychiatric testimony at issue was inadmissible on the question of guilt but admissible for purposes of determining punishment. See *id.* at 204.

110. *Id.* at 205 (Weintraub, C.J., concurring).

111. *Id.* at 206.

112. *Id.* at 207.

the notion of criminal responsibility, and he argues that our practices of placing criminal blame imply a rejection of that model. From this perspective, the argument for liability based on implicit bias seems to go directly against the grain of conventional arguments about the necessary conditions of holding someone responsible for an action.<sup>113</sup> Normally, to view an act as *caused* by factors outside the actor's control and awareness, and which are severed from the actor's conscious attitudes and beliefs, runs counter to holding the actor responsible.<sup>114</sup> The argument for liability in the "Work Experience" case is a direct rejection of that familiar philosophical dynamic. It is, in fact, an argument in favor of liability that depends on thinking of the employer's action in predominantly causal terms. The argument, counterintuitively, is that the employer should be held liable *because* she acts as a puppet controlled by her unconscious motivations. No wonder the idea of liability for unconscious discrimination seems so deeply controversial, even with all of the assumptions we have been indulging.

### C. IMPLICATIONS OF A COMMITMENT TO THE CAUSAL CONCEPTION

To say that the idea is controversial, of course, is not to say that it is wrong. My observations about the conceptual implications of liability for unconscious discrimination do not, in themselves, recommend for or against such liability. Perhaps we should *want* to shift our understanding of discrimination to accommodate a more scientific approach; perhaps we are misguided to conceive of human action as anything other than the result of causes that lie beyond our will.<sup>115</sup> At least as a purely textual matter, as I argued above,<sup>116</sup> Title VII's vague "because of" language seems to invite a causal interpretation of discrimination in a way that the fundamental principles governing criminal liability manifestly do not.<sup>117</sup> Yet, I do think there are important theoretical problems with a causal conception of discrimination that go beyond its merely being counterintuitive. I have explained that our felt reservation about liability for unconscious discrimination might be rooted in its connection with a causal conception of action. In the remainder of this Article, I offer some

---

113. I am making a claim here about precepts of choice and responsibility that inform our actual practices of holding persons responsible. *Cf.* SCANLON, *supra* note 104, at 277–80.

114. *See id.* Again, to be clear, I deny (for reasons elaborated by Scanlon) that responsibility presupposes control over the causes of our actions in any sense that would be incompatible with philosophical determinism. *See id.* For another concise argument that there is no relevant sense of "control" that governs attributions of responsibility, see MOORE, *supra* note 100, at 24–26.

115. For an example of this view based on psychological science, see DANIEL M. WEGNER, *THE ILLUSION OF CONSCIOUS WILL* (2002); *cf.* Deborah W. Denno, *Criminal Law in a Post-Freudian World*, 2005 U. ILL. L. REV. 601, 672–73.

116. *See supra* text accompanying notes 48–65.

117. My claim is about what the statutory text seems to permit, not about how courts have actually understood it.

possible concerns about incorporating such a conception into our legal understanding of discrimination.

*I. The Charge of Discrimination as Diagnosis*

The first worry is that the scientific, causal conception of discrimination diminishes the seriousness of the objection of discrimination as a moral criticism of an individual's or employer's action.<sup>118</sup> On the traditional, reasons-based conception, to object to an action as discrimination is to make a claim that the employer's rationale for the action was inadequate as a justification. The proscription of discrimination, so understood, has clear action-guiding, prescriptive force. From the perspective of the deliberating agent, it places a constraint on what the agent can adopt or endorse as a rationale for acting, a demand that the agent avoid acting on the basis of certain specified considerations. We are warranted in criticizing an agent who violates that demand, because everyone is morally answerable for acting on the basis of justifications that are institutionally illegitimate or that we could not reasonably expect others to accept.

The unconscious bias conception of discrimination shifts our focus from the deliberative perspective—the view of ourselves as agents who have the ability to choose the reasons on which we act—to a perspective in which our actions are the result of unconscious psychological forces that may be wholly unresponsive to any demand for reasons.<sup>119</sup> To what then, does the charge of discrimination on this view amount? Arguably, it ceases to have moral significance at all, at least as a criticism of the actor. In a scenario like that in the “Work Experience” case, where the actor can genuinely disavow any prejudicial motive, and where the actor's unconscious bias does not reflect her conscious attitudes, beliefs, and judgments, a charge of discrimination based on the fact of that bias is hardly a criticism at all of the actor-as-moral-agent.<sup>120</sup> The objection of discrimination takes on a diagnostic character rather than a moral one. It becomes primarily a claim that the action in question can be explained by

---

118. This concern is perhaps what underlies the argument advanced by certain critics that some constructions of implicit bias are too overinclusive to have real legal or political significance. See Arkes & Tetlock, *supra* note 8, at 264, 275. In the title of their critical article, Arkes and Tetlock ask the question, “Would Jesse Jackson ‘Fail’ the Implicit Association Test?” See *id.* at 257. The authors suggest that the answer is yes, which they take as a reason to be skeptical about the concept of implicit bias. See *id.* at 264, 270. Put more generally, the argument is that any test of discriminatory bias that could not be “passed” by someone like Jesse Jackson, an outspoken advocate for antidiscrimination and equality, could not be capturing any worthwhile conception of discrimination. For a response to this sort of argument, see Bagenstos, *supra* note 4, at 488–90.

119. See Denno, *supra* note 115, at 672–73.

120. For philosophical elaborations on this sort of view of the conditions under which it is appropriate to blame an agent for wrongdoing, see Pamela Hieronymi, *The Force and Fairness of Blame*, 18 PHIL. PERSP. 115 (2004); see also T.M. SCANLON, MORAL DIMENSIONS: PERMISSIBILITY, MEANING, BLAME 197–98 (2008); SCANLON, *supra* note 104, at 277–80; Angela M. Smith, *Control, Responsibility, and Moral Assessment*, 138 PHIL. STUD. 367 (2008).

reference to a certain psychological cause that intruded on the actor's active agency, not a criticism of the actor's exercise of that agency. The agent might embody or be a carrier of an inequality-producing bias. However, if that bias truly is not responsive to the agent's own judgments, then criticizing the agent for being influenced by that bias would be akin to criticizing an individual for infecting someone with a virus that the individual had no idea she was carrying.<sup>121</sup> Such criticism might be warranted to the extent that the individual had reason to take precautions against passing on the virus,<sup>122</sup> but absent such reason, the criticism would be misplaced.<sup>123</sup> The point is that if the charge of discrimination on the causal conception can, at least in some instances, entail nothing more than just such a diagnostic, explanatory claim, then that conception arguably weakens the moral significance of the charge of discrimination.<sup>124</sup>

## 2. *The Problem of Conceptual Instability: "Work Experience II"*

The second worry is that the causal-psychological conception of discrimination creates an instability in the law of employment discrimination. The instability can be introduced with a simple question: If we agree that an employer can be subject to liability on the basis of a *cause*, such as implicit bias, that lies outside the awareness and judgment-sensitive attitudes<sup>125</sup> of the decisionmaker, why should liability be limited to cases where that cause resides in the actor's *psychology*? What is so special about psychological causes of discrimination that lie in the actor? Return for a last time to the hypothetical employer in the "Work Experience" case. Let us change the facts a bit—call this variation of the case "Work Experience II." Suppose, now, that the employer does *not* suffer from any implicit racial biases. Imagine again that our hypothetical

121. Cf. Charles R. Lawrence III, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317, 321 (1987) (analogizing racism to a "disease" or "illness . . . that infects almost everyone").

122. This would be analogous to the unconsciously biased actor having reason to take precautions against the influence of her unknown biases.

123. For a rather different philosophical perspective, see GEORGE SHER, WHO KNEW?: RESPONSIBILITY WITHOUT AWARENESS (2009). On Sher's account, an agent can be morally responsible for an attitude that is causally connected to the agent's "constitutive features," see *id.* at 121–22, which need not be aspects of the agent with which he would necessarily identify. See *id.* at 122 (suggesting that even "neurophysiological mechanisms" might count as constitutive). Sher does not discuss the problem of unconscious bias in detail, however, and it is not clear to me whether his view implies that an agent will usually be morally responsible for her unconscious biases. The answer would depend, I suppose, on which aspects of the agent's psychology could be regarded as constitutive.

124. This worry about the causal conception of discrimination is not necessarily an argument against negligence-based or strict liability, either in the context of employment discrimination or elsewhere. Holding a person liable without requiring proof of intent is not the same thing as holding a person liable for an action *by virtue of* the action's being influenced by a causal factor outside the scope of the actor's agency. It is the latter sort of liability that is central to the causal conception.

125. SCANLON, *supra* note 104, at 20–24 (coining and explaining the term).

employer selects the white candidate over the black candidate for an open managerial position on the basis of the former candidate's superior work experience. But now let us suppose that we know a little bit more about the background of our hypothetical black candidate. It turns out that the reason that she has less work experience than the white candidate is simply that she had great difficulty obtaining employment in the early years of her career because of intentional discrimination by other potential employers. Absent such discrimination, the black candidate's work experience would have been at least as extensive as the white candidate's.

Is the employer in the "Work Experience II" hypothetical subject to liability for discrimination? I doubt anyone would say so. But if we have adopted a *causal* conception of discrimination that allows liability based on the influence of unconscious bias, it becomes strangely difficult to justify that answer. It is literally true, after all, that the employer's selection of the white candidate is, indirectly, "because of" the black candidate's race. The factor of race is quite clearly essential to a full causal explanation of the employer's decision. We can presume that if the candidate had been white, she would have achieved the same work experience as the white candidate and so would not have been rejected—at least not on the basis of work experience. We might be tempted to argue that the employer in the "Work Experience II" case had no control over the acts of prior potential employers and so cannot be held liable on that basis, but this will not do. The employer in the "Work Experience" case had no more control over her implicit and unconscious biases than the employer in the "Work Experience II" case had over the prior discriminatory causes of the black applicant's lesser experience. We might want to argue that the black candidate's relative lack of work experience is not the *fault* of the employer and so cannot be the basis for liability, but again, this will not do. The employer in the original case also was not necessarily at fault in any relevant sense for her implicit biases, yet those biases are the anchor of liability on the causal conception of discrimination. Finally, we might try to distinguish the cases on the grounds that the cause of the black candidate's rejection in the "Work Experience II" hypothetical is prior societal discrimination, rather than the employer's own biases, but that argument simply begs the question of why that distinction matters, given the irrelevance of the actor's awareness or responsibility under the causal conception of discrimination.

The point is that if, as is the case with the causal conception, liability for discrimination depends solely upon an inquiry into whether the aggrieved person's protected characteristics figures into an adequate causal explanation of the adverse employment act, it becomes very difficult to say why liability should be limited to those cases in which that

causal explanation goes *through the psychology* of the employer's decisionmaker. To put it another way, the black candidate's race is (by hypothesis) a necessary causal condition of the adverse employment action taken against her in both "Work Experience" and "Work Experience II". But if the employer's action in the original "Work Experience" constitutes objectionable discrimination simply in virtue of the causal influence of race in the history of that action, then the action in "Work Experience II" should also constitute objectionable discrimination. On the causal conception of discrimination, there can be no principled difference between the rejection of an applicant caused by the employer's unconscious racial bias, and the rejection of an applicant based on a job-related deficiency that was caused by past circumstances of racial inequality. Limiting liability to those adverse actions that are caused by *psychological* influences linked to the aggrieved party's protected characteristics is an unstable position, or in any event an analytically unsatisfying one.

### 3. *Biting the Bullets*

What are we to make of this instability? One possibility is that it sets up a *reductio ad absurdum* of the causal conception. If the causal conception of discrimination results in the absurd conclusion that the employer's action in the "Work Experience II" case constitutes objectionable discrimination, then that proves that the causal conception of discrimination should be rejected, as should the possibility of liability for unconscious discrimination, which depends on that conception.

There is, however, another important possibility. One could conceivably bite the bullet and abandon the notion that the psychology of the employer is of the essence in our legal conception of discrimination. In other words, one might argue that what I am calling an instability—the conflation, invited by the causal conception, of differential treatment caused by an employer's unconscious bias with differential treatment caused by an employer's reliance on facts in turn caused by societal racial inequality—actually points the way to a more adequate conception of discrimination. According to this "bullet-biting" view, we should accept that there truly *is* no meaningful difference between the "Work Experience" and *Work Experience II* hypotheticals. The phenomenon of implicit bias is nothing more than the embodiment in individual actors of the circumstances of inequality and patterns of racial disadvantage that continue to beset our society and our workplaces.<sup>126</sup> Implicit bias may operate through our individual psychologies, but it is a consequence and reflection of our social

---

126. See Lawrence, *supra* note 121, at 322–23. For a similar argument, see Justin D. Levinson, *Forgotten Racial Equality: Implicit Bias, Decisionmaking, and Misremembering*, 57 DUKE L.J. 345, 417–20 (2007).

history:<sup>127</sup> It “reflects the way people unknowingly carry society’s weaknesses with them at all times.”<sup>128</sup>

This seems to me a deeply plausible view. If we accept it, then the argument in favor of liability for unconscious discrimination truly is an argument for moving our conception of discrimination away from the paradigm of individual human action and more toward a conception that defines it as something like the expression of our persistent structures of social inequality.<sup>129</sup> Discrimination on this broader picture is not always the act of an individual agent who defies principles of equal treatment, but can also be the manifestation, through individual action, of past and persistent social injustice and inequality.<sup>130</sup> On this understanding of what discriminatory action is, holding employers liable for discrimination—both conscious and unconscious—must be understood not as a practice of enforcing norms of individual responsibility, but about effecting social change and reform.

### CONCLUSION

I set out in this Article to explore what might be called a naïve question about the legal significance of unconscious bias: Should employment discrimination law allow for the imposition of liability based on proof of such bias? To reach the merits of this question, I devised a thought experiment in which I hypothesized a set of facts (the “Work Experience” case) in an employment context and asked whether the employer should be held liable for discrimination if we could prove that her action was influenced by the operation of implicit bias, even if she genuinely believed that she was acting only on the basis of adequate reasons.

I have explained that the argument in favor of holding the employer liable in the “Work Experience” case on the basis of her implicit racial bias implies a shift in the operative model of discrimination from a justificatory conception (in which discrimination is centrally defined by a certain kind of inadequacy in an agent’s putative rationale in acting) to a causal conception (in which discrimination is defined by the presence of a certain kind of causal influence in an action’s psychological etiology). Our misgivings about liability for unconscious discrimination can be explained, I have argued, by its dependence on the causal conception.

---

127. See Levinson, *supra* note 126, at 417–20.

128. *Id.* at 420.

129. Cf. Lawrence, *supra* note 121, at 322–23 (arguing that an individual’s racism is intertwined with society’s belief system).

130. Larry Alexander makes a similar observation in discussing the difficulties of drawing a meaningful moral distinction between actions based on conscious bias, actions based on unconscious bias, and actions based on preferences that are facially neutral but are in fact “tainted” by a causal connection to historically discriminatory social structures. See Alexander, *supra* note 33, at 177–81.

Entertaining the normative plausibility of such liability calls forth the familiar philosophical tension between viewing an action as the determined consequence of antecedent causal conditions that are not responsive to the agent's own judgments, and at the same time holding the agent responsible for that action.

I have suggested that the causal conception of discrimination is susceptible to some worries—one having to do with diminishing the moral seriousness of the charge of discrimination, and the other with a potential instability that might undermine the distinction between individual and “societal” discrimination. What we ought to conclude from these worries is not entirely clear to me. It does seem, though, that if the model of implicit bias really is, as many commentators have argued,<sup>131</sup> the paradigm of discrimination that will be most relevant to the workplace in coming years, then perhaps it is not so naïve after all to ask whether the law should adjust accordingly. If this means changing our understanding of discrimination from an agent-centered, moralistic conception to a predominantly psychosocial, diagnostic one, then perhaps our wisest response might be to bite the necessary bullets and adopt that latter conception. If unconscious discrimination really is best characterized as akin to passing on an infectious disease, then maybe the law should approach the problem of such discrimination not in the traditional manner of assigning individual responsibility and blame, but much more in the manner of addressing an issue of public health.<sup>132</sup>

---

131. See, e.g., Blasi, *supra* note 21, at 3–16; Jolls, *supra* note 71, at 72–73; Krieger, *supra* note 2, at 1164.

132. See Lawrence, *supra* note 121, at 329.

\*\*\*