Taking Implicit Bias Seriously

Devon Carbado & Jerry Kang*

Version 1.9
January 07, 2024

Please do not quote, cite, copy, or distribute further, without explicit permission.

Thanks to: Sunita Patel, Noah Zatz, Jonathan Feingold, etc.

RAs: Justin Lee, Isabel Thompson

* Devon Carbado is the Honorable Harry Pregerson and Distinguished Professor of Law, UCLA. Jerry Kang is Distinguished Professor of Law and Asian American Studies and Founding Vice Chancellor for Equity, Diversity and Inclusion (2015-20), UCLA. <kang@law.ucla.edu> <http://jerrykang.net>.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>I. The Implicit Bias Story</td>
<td>6</td>
</tr>
<tr>
<td>II. The Left’s Concerns</td>
<td>10</td>
</tr>
<tr>
<td>A. The Primacy of Structure</td>
<td>11</td>
</tr>
<tr>
<td>B. The Perpetrator Perspective</td>
<td>15</td>
</tr>
<tr>
<td>III. The Structural Case For Implicit Bias</td>
<td>20</td>
</tr>
<tr>
<td>A. Teaching the Firm</td>
<td>26</td>
</tr>
<tr>
<td>1. The Quadrants of Responsibility</td>
<td>26</td>
</tr>
<tr>
<td>2. The Structural Turn</td>
<td>35</td>
</tr>
<tr>
<td>B. Teaching the Academy</td>
<td>36</td>
</tr>
<tr>
<td>Conclusion</td>
<td>44</td>
</tr>
</tbody>
</table>
INTRODUCTION

Although now a distant memory, recall that a few months before the 2020 Presidential election, Donald Trump issued an Executive Order preventing federal contractors or grantees to provide diversity trainings that conveyed “divisive concepts.”¹ Some of these concepts were caricatures of Critical Race Theory (CRT), which the Administration attacked by name.² But the Executive Order also targeted workplace trainings suggesting that “an individual, by virtue of his or her race or sex, is inherently racist, sexist, or oppressive, whether consciously or unconsciously . . . .”³ The deliberate use of “unconscious” explicitly targeted implicit bias. And, indeed, outside of the context of the Executive Order, then-President Trump made clear that he viewed implicit bias education as un-American.⁴

That the Right would target implicit bias was predictable. After all, the evidence of implicit bias undermines the self-flattering view that we are already “colorblind” and treat each other solely by the content of character, not the color of skin.⁵ Moreover, implicit bias makes it harder to insist that persistent racial inequality is simply a product of individual talents,

³ Id (emphasis added). The word “inherently” does tremendous work throughout the Executive Order, but it is nowhere defined.
⁴ In the interest of transparency, we have each served as the Faculty Director of the Critical Race Studies program at UCLA School of Law. We each have published and spoken on implicit bias. One of us has had a Senate confirmation blocked because of our worked in Critical Race Theory and implicit bias.
⁵ Jerry Kang & Kristin Lane, Seeing Through Colorblindness: Implicit Bias and the Law, 58 UCLA L. REV 465 (2010). MLK’s rhetoric is deployed in §1 of the E.O. For a general critique of the way in which colorblindness has been mobilized to legitimize racial inequality, see Neil Gotanda, A Critique of “Our Constitution is Colorblind”, 44 STAN L. REV. 1, 1-68 (1991).
efforts, and agency. The promulgation of the Executive Order was also nakedly political, as explained powerfully by two of our colleagues. Cheryl Harris has argued that the Executive Order was yet another attempt to manufacture a racialized bogeyman. And, according to Kimberlé Crenshaw, that racialized bogeyman metastasized into a broader campaign to cancel not only implicit bias and CRT, but also educational engagements around race more generally.

We agree with our colleagues’ critique. Indeed, with Florida Governor Ron DeSantis leading the charge, conservatives have launched an aggressive cancel culture campaign targeting anything that smacks of anti-racist education. All sorts of books are being banned, including those authored by James Baldwin, Toni Morrison, and Zora Neale Hurston; all sorts of ideas are being expunged from the curriculum, such as intersectionality, structural racism, and Black feminism; and all sorts of topics are being taken off the K-12 educational table, including those focused on reparations, mass incarceration, and the Movement for Black lives. The Right’s critique of implicit bias education is

---


part of this broader conservative cancel culture ecosystem, one that is likely to become even more salient during the upcoming 2024 electoral cycle.

Why has the Right specifically targeted implicit bias education for suppression? Part of the answer is that the Right genuinely fears the implications of taking implicit bias seriously. More precisely, they fear that implicit bias may recharacterize persistent racial disparities as evidence that racism is both pervasive\(^{10}\) and systemic\(^ {11}\) for which we can dodge neither individual nor collective responsibility. To put the point more sharply, the Right fears that implicit bias has racially transformative possibilities.

To some on the Left, this claim sounds silly because they doubt that implicit bias even speaks to systemic bias much less nurtures transformative possibilities. From that vantage point, implicit bias is woefully inadequate both diagnostically and prescriptively because it misdescribes the racial inequality landscape. On this view, implicit bias simply isn’t “structural” enough—and might not even be “structural” at all—and thus cannot account for various problems described as “systemic” or “structural racism” (for example, mass incarceration). Nor is implicit bias education useful for social movement actors who seek to fundamentally transform society (for example, by way of reparations).\(^ {12}\)

This Essay grapples with the Left critique.\(^ {13}\) We focus on the Left because we believe that teaching implicit bias can be a profoundly important antiracist practice. Of course, it is not


\(^{10}\) In this essay, we focus on race and racism although the arguments cross-apply to other forms of social category bias and inequality.

\(^ {11}\) We use “systemic” and “structural” interchangeably.

\(^{12}\) On recent example of this critique from the left appears in the Harvard Law Review by Charles Lawrence, one of the original and most prominent critical race theorists, a scholar whom I know and admire. In his review of the book *BIASED* by social psychologist Jennifer Eberhardt, Lawrence presents the absurdity of why we should discuss “Implicit Bias in the Age of Trump,” Charles Lawrence, 133 HARV. L. REV. 2304 (2020); see also JONATHAN KAHN, *RACE ON THE BRAIN: WHAT IMPLICIT BIAS GETS WRONG ABOUT THE STRUGGLE FOR RACIAL JUSTICE* 10-12 (2018); Samuel R. Bagenstos, *Implicit Bias’s Failure*, 39 BERKELEY J. EMP. & LAB. L. 37 (2018).

\(^ {13}\) Since any reference to “the Left” or “the Right” will be over- and underinclusive, please read these terms as rough markers on an ideological spectrum, not as monolithic ideological categories.
always or necessarily so. As with any topic, implicit bias can be
taught badly, for example by mischaracterizing the science, by
encouraging obsession over individual implicit bias scores, or by
suggesting that simplistic interventions taken exclusively at the
level of individual behavior will suffice.

But implicit bias can also be taught to motivate and guide
structural reform in both public and private spheres, in both
legal doctrine and institutional design, and in arenas in which
racial subordination has long been naturalized, such as policing.

Our broader claim is that in the current social and political
moment—fraught with racial divisions, contestations,
retrenchments, and debates about core features of our
democracy—it makes strategic (and not just normative) sense to
retain implicit bias education and argumentation in our
progressive toolkit. We are not saying implicit bias must figure
as the most important tool in that kit. For some it will, for
others it won’t, and we should embrace that diversity of approach.
Our modest suggestion is that by appreciating this tool’s
particular capacities and limitations, those broadly identifying
with the Left can better persuade institutions to experiment with
structural reforms that would not be seriously considered
otherwise. This Essay demonstrates how. Motivating this effort
is our intuition that if the Right is taking implicit bias
seriously for fear of its transformative possibilities, the Left
should take it seriously to realize them.

14 For an argument linking implicit bias to interventions within Critical
Race Theory, see Devon W. Carbado & Daria Roithmayr, Critical Race Theory
Meets Social Science, 10 ANNU. REV. L. SOC. SCI. 149 (2014).

15 See Devon W. Carbado, From Stopping Black People to Killing Black People,
105 CALIF. L. REV. 125 (2017) (drawing a nexus between implicit bias and
structural accounts of police violence); Jerry Kang, Judge Mark Bennett,
Devon W. Carbado, Pam Casey, Nilanjana Dasgupta, David Faigman, et al,
Implicit Bias in the Courtroom, 59 UCLA L. REV. 1124 (2012) (illustrating how
implicit bias can shed light on the racialized dimensions of the criminal
justice system).

16 See, e.g. Robin Walker Sterling, America’s Cycle of Racial Progress and
Retrenchment, WASHINGTON POST, November 6, 2022, at B3; Jamelle Bouie, It Is
a Well-Known Truth That Opponents of Democracy Don’t Want You to Have
Nice Things, N.Y. TIMES (Sept. 13, 2022),
https://www.nytimes.com/2022/09/13/opinion/democracy-inequality-united-
states.html; Suzanne Mettler & Robert C. Lieberman, Four Deadly Threats to
American Democracy are Raging All at Once, L.A. TIMES, Nov. 17, 2020, at
All.

17 A similar point can be made about Critical Race Theory: Which is to say,
it’s precisely the fact that the right is waging a war against this body of
We begin in Part I with the briefest of primers on implicit bias. The aim is to outline the core ideas and empirical foundation on which the literature of implicit bias rests.

Part II then foregrounds and interrogates the most salient dimensions of the Left’s critique. In doing so, we reveal where we agree and disagree with the Left’s concerns, introduce the concept of “racial sedimentation” to make clear that eliminating implicit bias would not magically eliminate racial inequality writ large, and reject a puzzling misunderstanding that implicit bias education privileges antidiscrimination models that require the showing of “purposeful intent.”

Part III shifts the discussion to highlight how taking implicit bias seriously has theoretical, practical, and legal purchase in facilitating structural interventions. First, theoretically, we focus on how implicit bias helps support and rearticulate central ideas in Critical Race Theory, a body of work that is committed to structural understandings of race.18 We pay particular attention to the claim that race is a social construction, which entails the attribution of social meanings to made-up racial categories. Our central point is that the evidentiary story the implicit bias literature tells about racial attitudes and stereotypes is, at bottom, a story about race and social meaning—namely, that social meanings play a role not only constructing race (for example, by advancing the idea that to be Black is to be racially inferior), but also facilitating and legitimizing racial inequality (for example, by serving as a rationalizing discourse for various forms of Black subjugation, such as slavery).

Second, practically, we explain how implicit bias education can encourage law firms to think structurally about race and law firm hiring. Our point of departure is a conceptual framework we call “quadrants of responsibility.” Informing that framework is the view that structural change requires individual behavior. Which is to say, because structural changes don’t happen by themselves, anti-racist proponents must persuade individual people, particularly those in positions of power, to advocate for or at least tolerate the experimentation with structural reforms.

work that the Left should defend it. See, e.g. What These Attacks Are Really About, Afr. Am. Pol'Y F., https://www.aapf.org/_files/ugd/b77e03_ba311e3f3aa84ca6aadbf73b32a3b289.pdf (last visited Sept. 26, 2023).

18 For an articulation of the core ideas that underwrite CRT, see Crenshaw et al, Introduction, Critical Race Theory: The Key Works that Formed the Movement. See also Devon W. Carbado, Critical What What.
But there are deep and profound challenges to doing so. In particular, and as our “quadrants of responsibility” illuminate, because many people think of racism as something that happened largely “back then” and/or is a phenomenon that isn’t about “me,” they are less likely to view racism as a phenomenon for which they should take personal responsibility. We describe how implicit bias education can teach the firm to overcome that challenge and invite those institutions to ask themselves whether they are living up to their professed normative commitments. Note, then, that implicit bias education typically does not demand that institutions adopt new norms, morals, or ethics and insist that they become more virtuous, charitable, or generous. Instead, the implicit bias gambit meets institutions where they are and simply makes an evidence-based case that they have failed to meet the minimum normative commitments that have already publicly embraced. In sum, implicit bias, anchored by the science, motivates many people not only to “sit up” and listen but also to think about what they can do—at the levels of both individual behavior and institutional/structural reform. We highlight the implications of that insight for effectuating changes at law firms, particularly in the domain of hiring.

Third, legally, we highlight how courts have begun taking implicit bias seriously to significant doctrinal effect. By appreciating the importance of implicit bias as potential mechanism for producing and instantiating racial discrimination, courts have redesigned various doctrines in ways that are more consistent with a structural approach.

I. THE IMPLICIT BIAS STORY

We begin with some basic definitions. For purposes of this story, a “bias” is deviation in one’s thinking from a presumptively neutral baseline. If, for example, the baseline presumption is that every student should get a computer, a teacher’s thought that students who live in “bad” neighborhoods should not because the equipment would be lost or damaged, would reflect a bias.

In terms of social cognitions, biases can show up in our “attitudes” and “stereotypes” about groups of people. An attitude is an overall evaluative valence (ranging from positive

---

19 When we discuss the science of implicit bias, we try to maintain a conceptual distinction between, on the one hand, “bias” (a social cognition in the form of stereotype or attitude) and, on the other, “behavior” (a manifestation of the bias in judgment, decision, reaction, or treatment). Such behavior may also amount to legally cognizable “discrimination.”
to negative) we have toward a particular social category (e.g., Black people). In lay terms, our attitudes track roughly whether we like or dislike that group or category. A stereotype, by contrast, is a generalization about a social category. We make them all the time because we naturally think through categories. For example, we stereotype old people as grumpy and young people as carefree. The key feature of stereotypes is that they probabilistically associate traits (e.g., grumpy) to groups (e.g., elderly).20

Importantly, stereotypes and attitudes can flow in different normative directions. For example, one might generally like African Americans (a positive attitude) but think they are intellectually inferior (a negative stereotype). Conversely, one might generally dislike Asian Americans (a negative attitude) but think they are mathematically superior (a so-called “positive” stereotype).21 The fact that stereotypes and attitudes of the same group can flow in different normative directions is one reason why formulations of “bias” should distinguish between attitudes and stereotypes.

Let’s now think about the core distinction between “explicit” and “implicit” forms of bias. A good place to start is knowing that when bias is “explicit,” it is subject to direct introspection and can be measured simply by asking ourselves for a truthful accounting. We can simply query ourselves about our attitudes and stereotypes about a social group and get the answer back. Our explicit biases are available to us because we are aware that we have them.

Of course, anyone who keeps a secret knows that private awareness and public disclosure are not the same thing. The fact we are able to gauge our explicit biases does not mean that we are always willing to share them openly. Because revealing biases can generate moral opprobrium and social sanction, we regularly hide or cover up biases we know we actually have, a phenomenon that we might call “covert biases.” By way of example, suppose

20 For a more extended articulation of these terms, see Jerry Kang, Trojan Horses of Race, 118 HARV. L. REV. 1489 (2004).

21 So-called “positive” stereotypes can reinscribe unflattering traits by negative implication. For example, technical superiority might simultaneously imply lack of creativity or charisma. Also, “positive” stereotypes can produce harm indirectly. If Asian Americans are lauded as the “model minority,” might that breed resentment from other racial minority groups that are stigmatized as inferior by comparison? Might the “model minority” stereotype blind us to the fact that some Asian Americans might be victims of hate crimes, looked over for promotions, and need a social safety net?
during a Freshmen orientation session, we asked students to raise their hand if they thought Asian Americans were on average better at math than African Americans. Or that African Americans were on average better athletes than Asian Americans. Chances are few, if any, students would raise their hand even if they privately believed the stereotypes to be accurate, again on average. This is an example of a covert bias, an explicit stereotype that is kept hidden or deliberately concealed.

Let's contrast our definition of “explicit bias” with a definition of “implicit bias.” To begin, analytically, “implicit” bias should not be mistaken for a concealed explicit bias. Instead, implicit biases are attitudes and stereotypes that people cannot readily ascertain through direct introspection. Some people may know about the phenomenon of implicit biases as a general matter, but even for these people, honestly asking themselves about their implicit biases won’t always generate a complete or accurate answer. Because implicit biases are mostly introspectively inaccessible, some other measurement device is necessary to capture them.

One family of instruments uses millisecond differences in reaction times, based on the idea that faster responses reveal stronger mental associations as compared to slower responses. The most extensively used and best validated example of such a measure is the Implicit Association Test (IAT). The IAT can measure, among other things, the relative speed at which we associate dominant groups (for example, heterosexuals) and minority groups (for example, gays and lesbians) with good and bad words (evincing a positive or negative attitude). When such implicit measures are taken broadly, across millions of people, throughout the globe, over decades, scientists have found that implicit biases exist, are pervasive, and generally larger in magnitude than the explicit biases we admit to on anonymous surveys.

Decades of social psychological research on attitudes and stereotypes suggest that explicit biases predict an individual’s judgment of and behavior toward the relevant social category to a small degree. In the past 20 years, implicit social cognition research has demonstrated the same for attitudes and stereotypes that are implicit. In fact, in behavioral domains that are politically or socially sensitive (such as interracial
interaction), implicit bias measures predict behavior slightly better than explicit measures.

We do not want to exaggerate. The correlations found in meta-analyses between the IAT and discriminatory behavior are small to moderate in effect size, ranging from Pearson’s $r = 0.10$ to 0.24. But as one of us has recently detailed, a small Pearson’s $r$ can raise a large social concern. Consider these examples.

Have you ever told your child not to smoke for fear of lung cancer? Have you ever used bottled or filtered water for fear of what low-level lead exposure might do to your child’s brain? Would you be shocked to find that the correlations between ever smoking and lung cancer in the next 25 years, or low-level lead exposure and reduced childhood IQ are lower than the correlation found between implicit bias and discriminatory behavior? And when numerous small effects are integrated over time (e.g., over an entire professional career) and across large numbers of people (e.g., all Whites versus people of color), they can produce surprisingly large consequences, especially in hypercompetitive tournaments.²⁷

That, in a nutshell, is the implicit bias story. As we will discuss in greater detail further along in this Essay, the fact that the phenomenon is “evidenced-based” and mostly received as such by politically diverse audiences is part of why implicit bias education is potentially powerful. We are not suggesting that genuine scientific disputes and uncertainties have all been resolved. But notwithstanding those disputes and uncertainties, we note that the evidence of implicit bias rests on surer scientific footing than our “commonsense” lay psychological intuitions about discrimination generally and how objective we are specifically.

Finally, in pushing for implicit bias education we do not naively presume that everyone will evaluate the evidence simply on their merits, especially given the Right’s investment in “manufacturing doubt” in ways that go beyond the ordinary skepticism inherent to good science. Instead, we are asking whether, from an antiracist perspective, we are better or worse off with the additional empirical payoff provided by implicit bias.

²⁷ For the full argument, see Jerry Kang, Little Things Matter a Lot: The Significance of Implicit Bias, Practically and Legally, Daedalus (American Academy of Arts & Sciences forthcoming 2024).
research. Our answer is “better off”—and we don’t think it’s a close call.

But some of the Left think we are wrong. Accordingly, let’s turn our attention to the Left’s concerns.

II. THE LEFT’S CONCERNS

According to the Left’s critique, implicit bias education invites us to misunderstand racial inequality as bias largely lodged in people’s heads. As Charles Lawrence recently put it in the Harvard Law Review,

this [implicit bias] worldview sees the problem of racism as a problem of brains that are wired to create implicit bias. . . . My own view is that the larger forces at work are the forces of structural and institutional racism and the narrative of white supremacy that justifies those structures and institutions.  

Worse, one could argue that implicit bias embraces the “perpetrator perspective,” which has long been decried within Critical Race Theory. The perpetrator perspective locates all the crucial questions about discrimination in the perpetrator’s state of mind. This privileging of the perpetrator perspective helps explain the logic of the Supreme Court’s decision in cases such as Washington v. Davis.

We respond with both agreement and disagreement. On the one hand, we fully agree about the significance, indeed the primacy, of structural racism (over, say, interpersonal bias) in shaping people’s lives. On the other hand, we disagree with the criticism that implicit bias education somehow adopts the perpetrator perspective and necessarily undermines structural accounts of racial inequality. As we explain below, implicit bias can and should be employed to articulate structural accounts of race and can and should be employed to challenge perpetrator-centered models of discrimination.

---


A. THE PRIMACY OF STRUCTURE

We unequivocally agree with the Left about the primacy of structure. Eliminating the implicit biases in our heads would not eliminate the racial inequality in our social world. To demonstrate, suppose that a magic wand could immediately alter all our implicit biases to line up perfectly with our explicit biases, with no dissociation in any person’s brain. Let’s call this a “post implicit bias society.” Would waving that wand somehow immediately change how tomorrow looks? No! It would not immediately change where we lived, what our neighborhoods look like, what our schools look like, what our bank accounts look like, what our jails and prisons look like, what our board rooms look like, where the highways bisect our cities, what media representation looks like, etc. This is at least one sense in which eliminating implicit biases would not, ipso facto, eliminate the material conditions of race.

One obvious reason those inequalities would persist is because explicit bias would remain. In other words, stone cold racists, dog-whistling trolls, and those confident about the accuracy of their racial stereotypes and the appropriateness of their attitudes would continue to act on them. A less obvious but more fundamental reason racial inequality would persist is because the disappearance of implicit bias—and even of explicit bias—would do little to disappear what we call “racial sedimentation.”

By “racial sedimentation” we mean the accumulation of racial advantages and disadvantages over time, including in the form of intergenerational family wealth and intergenerational economic precarity, respectively. The key to understanding racial sedimentation is recognizing that regimes of racial subordination steal opportunities not only in the present (for example, by denying people access to certain forms of employment and education). They also steal opportunities into the future (for example, by denying people the opportunity to develop human, social, and economic capital over time).

Consider the preceding points with respect to a specific historical example from 1954. In that year, the Supreme Court decided one of the most important cases in American constitutional history, Brown v. Board of Education, the case that ruled that “separate but equal” in the context of K-12 public education is unconstitutional.\(^{32}\) Sit with that decision for a moment. Imagine the sense of hope the opinion generated

especially in Black Americans, the sense of joy, the sense of possibility—and even the sense of freedom.

Now fast forward to 1955, one year after Brown was decided. What has changed? In some ways, an awful lot. Plessy v. Ferguson and its regime of de jure racial segregation is no longer a formal fixture of constitutional law: Separate is no longer equal. But it’s also fair to say that Brown left in place all the accumulated advantages and disadvantages up to that moment. If in the past Black parents were forced to attend segregated and resource-starved schools that necessarily contributed to worse employment trajectories and economic and social capital development, those realities would not fundamentally change in the future because their children could now legally attend integrated schools. Instead, the disadvantaged dimensions of Black parent’s lives (where they lived, how they lived, and the various forms of capital under-accumulated) would “carry over” into and remain a sedimented feature of their lives post-Brown. To put this another way, many people who were poor and had few economic opportunities pre-Brown didn’t escape that precarity post-Brown.

Figure 1 elaborates this point by cataloging various continuities between Black precarities and vulnerabilities Pre- and Post-Brown. The left column in Figure 1 identifies the social position African Americans occupied in U.S. society in the Plessy era. The list is not intended to be exhaustive. The right column then asks and answers—yes or no—whether that social position “carried over” into and thus was a structural feature of the newly-emerging post-Brown American society.

---

33 Plessy v. Ferguson, 163 U.S. 537 (1896).
<table>
<thead>
<tr>
<th>Black Marginality &amp; Vulnerability Pre-Brown</th>
<th>Was This Marginality &amp; Vulnerability “Carried Over” Post-Brown (Yes/No)?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic precarity</td>
<td>Yes</td>
</tr>
<tr>
<td>Politically disempowered</td>
<td>Yes</td>
</tr>
<tr>
<td>Underrepresented in colleges and universities</td>
<td>Yes</td>
</tr>
<tr>
<td>Underrepresented in Congress</td>
<td>Yes</td>
</tr>
<tr>
<td>Underrepresented on the judiciary</td>
<td>Yes</td>
</tr>
<tr>
<td>Vulnerable to racial violence</td>
<td>Yes</td>
</tr>
<tr>
<td>Vulnerable to employment segmentation (in lower paying jobs)</td>
<td>Yes</td>
</tr>
<tr>
<td>Vulnerable to employment discrimination and harassment</td>
<td>Yes</td>
</tr>
<tr>
<td>Vulnerable to premature death (because of both poverty and limited health care access)</td>
<td>Yes</td>
</tr>
<tr>
<td>Vulnerability to housing inequality</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Even a glance at Figure 1 makes clear that Brown did not undo the racially subordinating effects of Plessy (and slavery before it). That is because those subordinating effects had sedimented over time. They had created powerful inopportunity structures—circumscribed social and economic mobility possibilities—for Black people into the future. In that regard, Plessy ensured that Black people would be marginalized not only pre-Brown, but also post. Because “separate but equal” sedimented racial inequality, it was inevitable that institutions across various sectors of American life would play that inequality forward—almost “naturally”—into the world of Brown. Which is to say, with respect to the racial subordination Plessy effectuated, the opinion was sure to have an “afterlife” that structures the racial landscape to this day.

As we explain more fully elsewhere, one of the reasons for racial sedimentation is that antidiscrimination interventions in the United States have functioned largely as behavioral “injunctions,” not as monetary “damages.” For example, we formally ended slavery but did not redistribute the “unjust enrichment”—material and psychic—Whites acquired as a result of slavery. Nor did we compensate the survivors of slavery for the harms they suffered under that regime. The same point can be made about the formal ending of Jim Crow. The deconstitutionalization of “separate but equal” functioned as an injunction without redistribution or compensation. It did not dissolve the sedimentation of racial inequality. Thus, while Brown was instrumental in dismantling the formal edifice of Jim Crow, the opinion did very little to give back to Black people the economic, social, and political futures “separate but equal” had already stolen.

Our discussion of racial sedimentation should make clear that we are very much concerned about “structural racism.” Indeed, racial sedimentation invites us to take structural racism seriously by performing two important functions. First, racial sedimentation challenges the ubiquitous presentist narrative that frames the past as embarrassing but irrelevant in today’s
Second, racial sedimentation complicates the view that people's lot in society is a function of their individual talents and efforts, without regard to the disparate resources made available to them through the racialized transmission of intergenerational capital.

Racial sedimentation helps us understand why eliminating implicit bias going forward would not eliminate racial inequality writ large. In that sense, efforts directed exclusively at what might be called "implicit bias abolitionism" could not possibly be efficacious. The Left is entirely correct, then, to critique antiracist interventions that focus too much on ridding people of their implicit biases (or, more instrumentally, lowering their IAT scores). However, the fact that abolishing implicit bias would not abolish racial inequality is not the same thing as saying that implicit bias is a distraction or that implicit bias education cannot catalyze structural change. We disagree with the Left on that point and explain why below. But before getting to that point, we address another puzzling critique from the Left—that implicit bias education necessarily instantiates the "perpetrator perspective." We unequivocally reject that claim.

B. THE PERPETRATOR PERSPECTIVE

Most students of law will be familiar with the case Washington v. Davis—a core case in Constitutional Law, a course that virtually every law student takes. At issue in the case was the appropriate constitutional standard for determining an equal protection violation. The plaintiffs maintained that the Washington D.C. Police Department violated their right to equal protection by using a test that disparately impacted African American applicants.

The Supreme Court roundly rejected the plaintiff's claim. The Court reasoned that to sustain an equal protection challenge, a plaintiff must show that the alleged wrongdoer's adverse treatment of the plaintiff was purposefully and intentionally motivated. According to the Court, if there is no self-conscious purpose to harm people because of their race, then there is no cognizable equal protection problem, regardless of the disparate

36 This logic—that that past is the past—is a core feature of equal protection doctrine. See, e.g. Adarand v. Pena, 515 U.S. 200, 202 (1995) ("Under the Constitution there can be no such thing as either a creditor or a debtor race. We are just one race in the eyes of government.")


38 Id. at 232-233.

39 Id. at 238.
The Court was careful to point out that disparate impact could be evidence from which to infer a purposeful discriminatory intent. But disparate impact, standing alone, would not give rise to an equal protection violation.

Davis is widely criticized, certainly by the Left. It is also offered as Exhibit A for why implicit bias education is misguided or harmful. By focusing on individual-level cognitions, implicit bias scholars are criticized for embracing precisely the perpetrator perspective on which Washington v. Davis rests. And that perspective leaves us blind to even the most grotesque manifestations of disparate impact.

Unfortunately, this criticism fails to distinguish two separate dimensions of the perpetrator’s purposeful intent. The first dimension emphasizes the “perpetrator”—that is, the single actor accused of doing something wrong. But the second

---

40 Id. at 246.
41 Id. at 242.
42 Id.
44 “the behavioralist approach ‘accepts the central premise of the Davis intent requirement — that the harm of race discrimination lies in individual acts infected by bias’.” Lawrence 2020 tan [text accompanying note] 80. Jonathan Kahn makes a similar point in his book Race on the Brain. 58, 94. “debiasing thus boils down to an adjunct to the diversity rationale, hopped up by references to cognitive psychology and neuroscience but ultimately reinforcing the conservative move away from examining the harm visited upon the subject of discrimination with the mindset or the intent of the perpetrators of such discrimination.” Kahn, supra note 12, at 93.
45 Students of employment discrimination recognize that this is not always so for statutory frameworks. For instance, Title VII does accept a theory of “disparate impact” that can find a statutory civil rights violation even when no self-aware purpose is demonstrated. But, for the moment, let’s remain on the Constitution’s equal protection clause. In that important and rarefied realm, “intent” is key.
dimension emphasizes “purposeful” intent—that is the mental state that motivated or caused the perpetrator to act. On the one hand, implicit bias education does indeed typically focus on the individual perceiver or perpetrator (the first dimension) because implicit bias is primarily an individual-level measure (although, as discussed below, can be subsequently averaged over larger groups and geographies). On the other hand, implicit bias education explicitly rejects the notion that self-conscious, purposeful intent is strictly required for there to be cognizable harm (the second dimension). In other words, focusing on the individual does not mean endorsing the primacy or necessity of purposeful intent.

The criticism that implicit bias education acquiesces to the self-conscious, purposeful intent framework of Washington v. Davis is curious. After all, the very purpose of the implicit bias research is to pivot away from the dominance of purposeful intent. A central tenet of implicit bias education is that even if we lack any purposeful intent to harm people because of our explicit racial biases, it is nevertheless likely that implicit biases will causally warp our decisions unless we take precautions. Thus, behavioral realists—who insist that law should assimilate more accurate models of human behavior including findings about implicit bias—argue that the core of racial discrimination should require racial causation not purposeful intent.

That is precisely why all the implicit bias scholars we know lament Washington v. Davis for being

---

46 See discussion infra.

47 We see prominent examples of such nuanced thinking throughout the law. In both tort law and criminal law, for instance, we focus on individual-level behavior (i.e. individual tortfeasors or defendants) but tease out gradients of responsibility predicated on mental states less culpable than purposeful intent, such as “recklessness” and “negligence.”


49 There is a broader point about causation that it’s fair to say that the conception of implicit bias does not capture. It is entirely possible for a decisionmaker to promulgate a policy that (a) causes Black people to be excluded for a hiring pool, but (b) is not a function of implicit bias. Scholars sometime frames this scenario as a problem of disparate impact.
behaviorally unrealistic by limiting constitutional concern to only those behaviors that are purposefully intended.\textsuperscript{50}

A racial causation approach to antidiscrimination law would take seriously concerns about disparate impact. Consider this point with respect to \textit{Washington v. Davis} itself. In that case, the test the Washington Police Department used to hire police officers had a disparate impact on Black applicants. Under a causation model, the question is not whether the police department’s leadership who employed that test intended to discriminate against Black people. The question, instead, is whether the policy they promulgated “caused” that result. Moreover, replying in the affirmative wouldn’t necessarily sustain an equal protection claim. In fact, part of what is frustrating about the majority opinion in \textit{Washington v. Davis} is that it presents the doctrinal debate as one about whether an intent standard should govern or a disparate impact one. But one could easily proffer a doctrinal arrangement on which the establishment of disparate impact shifts the burden to the government to justify the use of the test.

Indeed the Supreme Court just four years earlier had embraced such an approach in the context of Title VII in \textit{Griggs v. Duke Power} (1971).\textsuperscript{51} In other words, the Court in \textit{Washington v. Davis} could have embraced the \textit{Griggs} three step process of (1) requiring the plaintiff to make a prima facie case of disparate impact; (2) next, requiring the defendant to explain the business necessity of the test; and (3) allowing the plaintiff to point out another practice with less disparate impact that could work just as well.

In the alternative, the Court could have required different levels of means-ends scrutiny. Some would undoubtedly call for “strict scrutiny”—meaning, the government would need a compelling justification for the Test 21 and the test would need to be “narrowly tailored” to advance that interest. Others might believe that strict scrutiny is too heavy a burden for the government to meet and something akin to “intermediate scrutiny” would be more appropriate. Still others might insist on something closer to “rational basis.” The point is that the establishment of disparate impact could operate as a burden-


\textsuperscript{51} The Court of Appeals had embraced just such an approach when it imported the Title VII statutory disparate impact theory into its equal protection analysis. 1512 F.2d 956, 958 n.2 (1975).
shifting device within the cause of action, not the establishment of the cause of action itself.

To avoid confusion, we are not saying that implicit bias explains all forms of disparate impact. Presumably it does not. To return to Washington v. Davis, we are not suggesting that, but for implicit bias, Test 21 would have been neither devised nor deployed. We just don’t know. But two points deserve emphasis. The first is that precisely because of the pervasive nature of implicit bias, one can almost always make the argument that implicit bias is plausibly at play, even if causal mechanisms distinct from implicit biases are also operating. Second, even assuming that the initial promulgation of Test 21 was not caused by implicit bias, the phenomenon could nevertheless be doing work. Concretely, implicit biases about African Americans as both intellectually deficient and criminally oriented, biases that would render Black people unqualified to be police officers, could help to explain the Washington D.C. Police Department’s decision to continue to using Test 21, notwithstanding the impact the test was having on African Americans.

The broader points we are making are (a) implicit bias is undoubtedly implicated in some (though not all) instances of racially disparate impact, including those that bear on employment, and (2) evidence of implicit bias encourages us to think about causation beyond inquiries into a “perpetrator’s” purposeful intent.

Implicit bias education pushes back against the perpetrator perspective in another, indirect way—by addressing the problem of “covert bias.” Here, too, Washington v. Davis is instructive.

Thinking beyond the Davis case, the scientific evidence that implicit biases are pervasive raises more seriously the possibility that people might be underreporting the explicit biases they know that they have. Empirical evidence exists to support this point.

This is not to argue that the IAT should be employed as a sort of polygraph to ferret out “perpetrators”—that is, people who evidence precisely the kind of conscious and purposeful racial motivations that Washington v. Davis requires. Indeed, we have argued against using the IAT in that way.55 The point is that just

because "covert bias" is hard to detect (people who harbor conscious racial animus or stereotypes typically don’t publicly declare themselves) does not mean that it no longer exists. And the overwhelming evidence of implicit biases is a useful and evidence-based reminder of this likelihood: Chances are, people harbor more negative views about traditionally racially marginalized groups than they openly disclose.

To summarize, implicit bias exposes two separate but related challenges that require us to move beyond purposeful intent: a "bias unawareness" problem (the fact that our behavior is shaped by biases we genuinely don't know we have) and a "covert bias" problem (the fact that our behavior is shaped by biases we knowingly conceal and keep covert). By illuminating both problems, implicit bias education undermines deference to the subjective, often-self-serving claims of the perpetrator privileged in Washington v. Davis.

III. THE STRUCTURAL CASE FOR IMPLICIT BIAS

This Part circles back to our agreement with the Left critique that dismantling structural racism should be at the forefront of anti-racist practices. Our goal is to foreground how implicit bias education can advance that project in ways that contradicts the Left’s critique that implicit bias education shifts the focus away from institutional and structural reform. We begin with a focus on theory and explain how taking implicit bias seriously within Critical Race Theory can meaningfully advance structural sensibilities about race.

A. THEORY: REARTICULATING CRITICAL RACE THEORY

Almost every structural account of race in the academy, especially within Critical Race Theory, begins with the assertion that "race is a social construction." While there are competing explanations of what that means, every articulation includes some version of the claim that racial subordination involves attributing particular social meanings to different categories of racialized bodies. To make this less abstract, think about the social meaning of people of Asian ancestry as disloyal and inassimilable. (In social cognition terms, these are stereotypes of Asian people.) There’s a direct relationship between that social meaning and the internment of people of Japanese descent during World War II. In other words, the structural features of the internment apparatus (e.g., curfew, evacuation, and detention

---


Do not cite or distribute in any way without the express permission of the authors.
to use the stylized terms of the Supreme Court) and their structural consequences (e.g., geographical dispersion, property escheats, and inter-generational disruption of wealth) rested upon a socially constructed view of an entire racial group.

We can tell a similar story about Chinese Exclusion by drawing attention not only to the infamous Chinese Exclusion Acts, which significantly circumscribed the terms on which Chinese people could enter the United States, but also naturalization legislation and caselaw, which prohibited people of Asian ancestry from naturalizing as citizens. These regimes, too, were predicated on racial assumptions about people of Asian descent as disloyal and inassimilable (stereotypes) as well as a general negative attitudes toward Asian people.

How does the modern-day scientific construct of implicit bias connect to these historical examples? To make explicit what we’ve been suggesting all along, racial bias (including attitudes and stereotypes) is arguably just another way of talking about the social meaning of race. Racial stereotypes, for example, encapsulate racial social meanings. The views that people of Asian ancestry are disloyal and inassimilable are group generalizations—i.e., stereotypes. Racial attitudes also encapsulate social meanings. The fact that there’s general positivity or warmth toward Whites versus negativity or chill toward Asians helps us understand racial subordination.

Another way to draw a nexus between the science of implicit bias, social meanings of race, and history is to reenlist the concept of racial sedimentation. Imagine, for example, that a progressive argues that the social meanings of race produce sediments that persist and build over time. Suppose that a conservative responds: “That’s plausible, but where’s the proof? After all, internment ended more than seventy years ago and people of Japanese descent are model minorities today. Why would we think that the racialized ideas that justified internment have an ‘afterlife’ in the contemporary moment, especially if most folks publicly claim that they’re not racist?” One answer—not the only and for many not even most important one—is that current measures of implicit bias tell us so.

On average, we still implicitly view Asians as more “foreign” than American as compared to Whites. That association, which had
been slowly decreasing, actually got stronger after March 2020, as terms like “Chinese virus” and “Kung Flu” circulated around social media and conservative news outlets.\(^{61}\) In addition, various studies show that this implicit stereotype of foreignness is correlated with judgments of Asian Americans, including whether they should be hired in jobs that pose national security concerns.\(^{62}\) We believe that such empirical evidence, which uses standard null-hypothesis testing, helps substantiate the structural argument that social meanings of race travel through and sediment over time.

Stay with us for one more example, this time with reference to African Americans and mass incarceration. The racial stereotypes of African Americans as criminal and dangerous are social meanings of Blackness.\(^{63}\) Those social meanings not only facilitate the incarceration of Black people (“Black people are dangerous so they must be incarcerated”), they also help to legitimize that racialized social arrangement (“The fact that Black people are incarcerated must mean that they are dangerous.”).

Against the background currency of the claim that we are all already colorblind and disavowals of racism as a cause of mass incarceration, discovering our implicit biases against Black people can be a provocative and productive educational tool. The point of that education would not be to demonstrate that a specific individual police officer, or a specific individual prosecutor, or a specific individual judge evidenced implicit racial bias in a specific individual case, but rather to suggest that, structurally, our average collective cultural consciousness of African Americans—the social meaning we attribute to that group writ large—includes the perception of them as criminal and dangerous.

This is a good place to link more explicitly our intervention to ideas that have been advanced in Critical Race Theory. We are thinking particularly about Charles Lawrence’s groundbreaking article, The Id, the Ego, and Equal Protection: Reckoning with

---


\(^{63}\) See Oliver, supra note 48.
Unconscious Racism. After describing what he perceives to be the profound problems with a purposeful/intent-based understanding of racism, Professor Lawrence proposes what he calls a "cultural meaning" test that could help guide judicial determinations of racism. For Professor Lawrence, "the cultural meaning of racial texts remains, for me, the central and most important idea in the article." In that regard, unpacking what Professor Lawrence means by "cultural meaning" is all the more important. According to Professor Lawrence:

I propose a test that would look to the "cultural meaning" of an allegedly racially discriminatory act as the best available analogue for and evidence of the collective unconscious that we cannot observe directly. This test would evaluate governmental conduct to see if it conveys a symbolic message to which the culture attaches racial significance. The court would analyze governmental behavior much like a cultural anthropologist might: by considering evidence regarding the historical and social context in which the decision was made and effectuated. If the court determined by a preponderance of the evidence that a significant portion of the population thinks of the governmental action in racial terms, then it would presume that socially shared, unconscious racial attitudes made evident by the action's meaning had influenced the decisionmakers. As a result, it would apply heightened scrutiny.

The unconscious racial attitudes of individuals manifest themselves in the cultural meaning that society gives their actions in the following way: In a society that no longer condones overt racist attitudes and behavior, many of these attitudes will be repressed and prevented from reaching awareness in an undisguised form. But as psychologists have found, repressed wishes, fears, anger, and aggression continue to seek expression, most often by attaching themselves to certain symbols in the external world. Repressed feelings and attitudes that are commonly experienced are likely to find common symbols

64 Charles Lawrence, The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism, 39 STAN. L. REV. 317 (1986).
65 Id. at 355-356.
particularly fruitful or productive as a vehicle for their expression. Thus, certain actions, words, or signs may take on meaning within a particular culture as a result of the collective use of those actions, words, or signs to represent or express shared but repressed attitudes. The process is cyclical: The expression of shared attitudes through certain symbols gives those symbols cultural meaning, and once a symbol becomes an enduring part of the culture, it in turn becomes the most natural vehicle for the expression of those attitudes and feelings that caused it to become an identifiable part of the culture. 67

We quote Professor Lawrence at length because we think that objective evidence of implicit bias is powerful support of his claims regarding “repressed feelings and attitudes.” Implicit bias provides one concrete measure of the racial significance “society” attaches both to particular dimensions of social life (such as racial segregation) and to particular laws, policies, and practices (such as stop and frisk). Thus, when people say that it’s nonsense to argue that racism is deeply implicated in mass incarceration, one can mobilize the scientific evidence of implicit bias—specifically, the pervasiveness of the implicit stereotype that associates Black faces with pictures of weapons, and the evidence that such implicit stereotypes are likely to influence the judgment calls of police officers, prosecutors, defense attorneys, judges, and probation officers.

To be clear: our claim that a unique benefit of implicit bias discourse is its evidence-based, scientific orientation does not require us to privilege “empiricism” or “science” over other ways of knowing. We need not take the position that empirical and scientific approaches to knowledge production are per se better than other intellectual approaches. Nor, in advocating for implicit bias, are we meaning to obscure the insidious ways in which racism and science has gone hand in hand, 70 a point that can (and should) be made about all the conventional disciplines, including those (like the humanities) that are perceived to be safer havens for engagements around race but whose disciplinary

67 Lawrence, supra note 59 at 355-356.

70 See Devon W. Carbado, Yellow By Law 97 CALIF. L. REV. 633 (2009) (discussing some of the ways in which science has been a racial project).
histories and contemporary intellectual practices betray problematic racial investments.  

Our point, instead, is both pragmatic and epistemological. The pragmatic point is that while some people are more moved by qualitative narratives, personal experiences, and intuitive flashes of recognition, others are more moved by quantitative measures, experiments, and null hypothesis testing. Consider, for example, the very different reasons why people agreed to take (or refuse) a vaccine for COVID-19. Some people were moved by peer-reviewed studies; others were moved by peers; still others were moved by tweets. Against the background of people’s differing sensibilities about what counts as persuasive evidence, it makes little sense for the Left to limit itself to one antiracist register. For pragmatic reasons, the left should embrace a diversity of approaches.

As for the epistemological point, we believe that our understanding of social phenomena is enhanced by examining them from multiple (and inter-) disciplinary frameworks. An analogy to climate change might help. Think about how preposterous it would be for an economist to claim that there are only two ways to understand climate change: (1) economics v. (2) physics. And by engaging in physics, including computational modeling of aerosols, we are detracting attention from the larger problem addressed by the economics of fossil fuel production, factory farming, and cap-and-trade markets. In response, a physicist would scratch their head, mostly because they would not understand why studying climate change through physics prevents studying climate change through chemistry and biology, much less economics. Instead, they would suggest that the choice is better framed as between (1) economics only v. economics plus every other discipline, operating at every other level of abstraction or unit of analysis. In our view, the same reaction is appropriate here, as we grapple with comparably wicked and vexing problem: racial injustice. Studying implicit bias doesn’t require us to give up other disciplinary approaches to engaging race.

We view knowledge production the way most universities do—as a “big tent” enterprise. With respect to addressing racism and racial inequality specifically, there is no good reason to leave any particular discipline behind. In that regard, we think it’s a

---

good thing that from history, to sociology, to statistics, to literature, to economics, among other fields of study, we see scholars mobilizing their disciplines to stage various forms of racial interventions.⁴ Those interventions embed a lesson for legal academics on the Left: that their antiracist theorizing and practices should be disciplinarily diverse and create space for incorporating implicit bias education into legal analysis, organizational reform, and public discourses about race, law, and inequality.

B. PRACTICE: REDESIGNING THE LAW FIRM

Now we move away from the heady theory of Critical Race Theory to the practical world of persuading law firms to create a more equitable working environment through structural reform. We believe that implicit bias education can play a crucial role in this project. As a preliminary matter, we concede that implicit bias education comes in various pedagogical forms, some of which can undermine, not support, structural accounts of racial inequality. As we have already said, troubling from our vantage point is an over-emphasis on individual-level “de-biasing” techniques, where the goal is to try to change people’s individual implicit bias scores. But there is no inherent reason why implicit bias education should, in staging interventions, focus exclusively on individuals. Nor is it the case that focusing on individuals necessarily entails abandoning structural and institutional reform initiatives. We elaborate below employing our “quadrants of responsibility” as a point of departure. As we will show, the “quadrants of responsibility” is useful pedagogical device for explaining why implicit bias education might be particularly powerful in the context of law firm governance.

1. The Quadrants of Responsibility

The Left’s critique of implicit bias reflects the view that any focus on the individual is presumptively problematic and misdirected. We have already challenged one version of that claim—that implicit bias education necessarily acquiesces in purposeful intent-based models of discrimination, such as in Washington v. Davis. We now make a related but novel point: Implicit bias education can motivate individuals to experiment with structural reforms. We advance our case by introducing what we call the “quadrants of responsibility.”
We believe that the degree of responsibility felt for racial injustice depends on two variables--“who” and “when.” In the diagram below, the horizontal axis represents “who” and the vertical axis represents “when.” Here are the basic contours:

**Quadrants of Responsibility**

Let’s start with the horizontal axis. If I committed the bad act, I will feel greater responsibility. By contrast if They did so, not so much. This “I” and “They” shapes contemporary discourses about race in numerous ways. Consider, for example, debates about reparations. One argument against any form of reparations to African Americans is the sense that “I” didn’t enslave or Jim Crow anyone. “They” did. Therefore, why should “I” have to pay for what “They” did?

“I” and “They” logics also surface in affirmative action discourse. Here, the question becomes: Why should “I” experience so-called “reverse discrimination” to advance the goal of diversity? “I” did not create the underrepresentation of African
Americans in colleges and universities. "They" did. The "They" in this instance sometimes refers to African Americans themselves (on the view that African Americans have not worked hard enough to ensure their academic success) and sometimes it refers to other White people who have mistreated Black people.

Of course, there are many intermediate points between the extremes of "I" and "They." For example, I may feel some group responsibility as a function of whether my family, my team, my nation, my profession, my gender, my race, etc. committed the act. The point is that the closer the connection between me and the actors responsible, the more the event will appear on the left side of the diagram.

As for the vertical axis of *when*, if the bad act happened recently, we tend to feel greater responsibility. By contrast, if the bad act happened a long time ago, indeed centuries ago, the feelings of responsibility dissipate. It's as if bad acts have a built-in half-life or statute of limitations. Think, for example, about the 2020 murder of George Floyd. That violent tragedy created a sense of urgency. Institutions of almost every sort across the United felt that they had to do something. But by 2024, that sense of urgency has already dissipated. Seemingly, even the most salient example of the disposability of Black life has a short shelf life.

Let’s now combine the X and Y axes by considering terms that subsequent to the #BLM moment routinely circulate in public discourse: conquest, slavery, Jim Crow. These horrific histories resonate powerfully within our personal moral framework. However, many people find it easy to quarantine these concepts into the upper right quadrant of the diagram. *They* did it, back *then*. Our quadrants of responsibility now look like this:
The point we are emphasizing is that conquest, slavery, Jim Crow, and internment might create a vague, general responsibility on the nation state, but for many, they do not create a sense of specific personal or firm-level responsibility because, to repeat, they did it, back then. (To avoid misunderstanding, we repeat that we are not endorsing this they/back then dodging of responsibility. We are merely describing the tendencies we see in public discourses about race).

Even when we switch focus to more recent horrors, such as White supremacists marching in Charlottesville screaming “blood and soil” and “Jews must die,” or high-profile police killings of Black Americans that beggar belief, it is again easy for people to cabin those atrocities, this time in the lower right quadrant. Although these activities are taking place roughly now, They (evil people) are doing it. “I” (a good person) am not. Thus, while we might reject those actions and even protest against them, we do not view such acts as ones for which we are personally
responsible. That is why our quadrants of responsibility now look like this:

```
+----------------+----------------+
| THEN           | NOW             |
| the WHEN axis  |                 |
+----------------+----------------+
```

We can continue to build out our quadrants of responsibility by focusing on the left-hand side of the diagram. One way to do that is to surface mistakes that we’ve made. Consider off-color jokes or inappropriate, demeaning, or offensive language people may have casually used back in high school that now sound deeply inappropriate. Or consider the donning of Blackface to attend a college Halloween party. Today, we may be genuinely and deeply embarrassed that we engaged in those troubling forms of conduct. But those mistakes can be fenced off as youthful indiscretion. As such, they reside in the upper left quadrant in the following way:
In short, in the upper-left quadrant, problematic utterances and self-presentations occurred back *then*, but this is *now*. We’ve learned and moved on, for the better.\textsuperscript{32}

This leaves the final quadrant in the lower left, which is where most people feel the core of personal responsibility. Central to this quadrant is the idea that if we did something wrong right *now* (or recently), we will—and it’s perfectly appropriate for the law to make us—take personal responsibility. The problem is that people’s self-perception in this quadrant can get in the way of them doing just that.

Consider this point with respect to people who are politically on the Left, the Center, and the Right. Although we are generalizing, there are specific rationalizing discourses for each group—across the ideological spectrum—that are based on
exculpatory self-perceptions. Those perceptions effectively function to dodge responsibility.

Let’s begin with the Left. The self-perception here is best captured by the rhetorical “I get it.” In other words, people on the Left perceive that they have sophisticated knowledge not only about race and racism but also about how to be an antiracist. From that vantage point, “I” am not the problem. The people who don’t get it are. Decontextualizing the origins of the term “woke,” the Right routinely describes this “I get it but you don’t” sensibility disparagingly as “wokeism.”

Shifting now to people in the Center. Their self-perception is that they are “colorblind.” For them, that means that they judge people only on their merits and not on attributes that are irrelevant to the decision, such as race. (In addition, they may have a mild commitment to race conscious affirmative action, viewing it as a cost society should accepts to advance diversity.) This commitment to colorblindness can also (wittingly or not) function as absolution. It is a way of saying: “Whatever the problems of race might be, as bearers of colorblindness, we are not to blame.”

People on the Right also contend that they are colorblind, but that commitment carries with it two important provisos. One is that it’s “rational” to act on the basis of accurate stereotypes; doing so should not be treated as taking race into account. Plus, for many people on the Right, it’s reasonable for policymakers to take into account what they perceive to be the “cultural differences” among racial groups. For the Right, those “cultural differences” explain why some racial groups (like Asian Americans) take advantage of the “American Dream” in ways that other racial groups (like African Americans) do not. Accordingly, the Right routinely invokes culture to shape various public policy debates, including the “culture of poverty” (to argue against robust social safety nets for the poor (especially
Black people)), “model minority” (to argue against affirmative action (especially for Black people)), and “Black-on-Black crime” (to argue for aggressive policing practices to support the “war on crime” and the “war on drugs” (especially against Black people)). The broader point is that people on the Right generally profess a commitment to colorblindness (notwithstanding their carveout for relying on race along the lines we have discussed), and they invoke that commitment to absolve themselves of any responsibility to address what even they concede to be significant racial disparities in the United States.

Filling in all the self-perceptions of the Left, Center, and Right (again, remembering that we are crudely simplifying), the lower left quadrant now looks like this:

**Quadrants of Responsibility**

```
THEN

the WHEN axis

NOW

Youthful mistakes
  using the "N" word
  Halloween Blackface

conquest
  slavery
  Jim Crow
  internment

* I’m rational
* I’m colorblind
  * I get it

They’re neo-Nazis
They’re racist cops

I = the WHO axis

THEY
```

***
Take another moment to examine the entire image and the quadrants that constitute it. The key takeaway is that the rationalizing discourses that occupy each of the quadrants diminish the likelihood that people will see themselves as being implicated in, or as bearing some responsibility for, extant forms of racial inequality. Precisely how much these discourses absolve the feeling of responsibility is hard to quantify. But common sense teaches that the more we view social problems as about “them”/not “us,” and as about “then”/not “now,” the less likely we are to see ourselves as part of the problem.

2. The Value of Implicit Bias

Suppose, then, that with respect to engaging a law firm one insisted on privileging the top left, the top right, and the bottom right quadrants, to encourage the firm to adopt more equitable hiring, work allocation, mentoring, and promotion practices. That project is likely to have limited traction because, as we explained earlier, none of those three quadrants trigger robust feelings of personal responsibility.

What, then, about an approach that goes hard at the lower left quadrant and directly accuses people and their institutions for being “racist” right now. Is this effort likely to be successful? Again, we think the answer is “no.” For one thing, without smoking gun evidence to back up the claim of racism—evidence that tracks the kind of showing the Court insisted upon in Washington v. Davis—the audience will likely get defensive, angry even, and reject those claims as unfounded and inflammatory. And if in the context of that anger and defensiveness managers are forced to adopt policies, procedures, and practices that constrain their expert discretion, the outcome is not likely to be good. The best-case scenario is that managers will be annoyed and grudgingly comply. The more realistic outcome is resistance, backlash, and retrenchment. Specifically, managers would adopt merely cosmetic compliance or engage in circumvention.

By contrast, suppose we are permitted to expand our arsenal of arguments to tell a more nuanced story that includes implicit bias? By educating the law firm of basic social cognition, we can introduce new facts proven up by substantial scientific evidence that demonstrate that we are neither entirely rational nor colorblind. This dismantles the self-serving views we currently enjoy. In the lower left quadrant, we can force a realization that there remains a gap between what we claim and who we are. This realization can then trigger a powerful personal responsibility
because the bias no longer can be said to be somewhere out there, taking place back then. Instead it is in our own house, in our own mind.

To make things a little more concrete, recall our brief discussion of the IAT. You take the test; you get a score. Often an embarrassing one. And all of this happens right “now.” Upon getting that score, one also learns the average scores of millions of people, from across the world, which helps to drive home an observation we made earlier—namely, that the phenomenon of implicit bias is pervasive. More to the point, against the backdrop of the massive amounts of data on implicit bias, it’s hard to hold on to the view that race and racism are problems that reside in some distant “back then.” Those data reflect a story about contemporary society “now.”

Importantly, our implicit bias education to law firms leverages studies that are specially relevant to that domain. For example, [Kang study]. Consider as well [Reeves study]. We are always careful not to overstate the empiricism on which these studies are based. At the same time, we emphasize the ways in which small headwinds—for example, repeatedly missing out on informal mentoring, constructive feedback, and career-stretching work—can have big effects over time, especially in hypercompetitive industries such as law.

Our experience has been that a thoughtful presentation of the implicit bias literature to a range of constituencies—judges, community organizers, the media, academics, lay people, police officers, in addition to law firms—gets people to pay attention and engage with the possibility that racism is a contemporary feature of social life, not a bygone social phenomenon. We’re not saying that we get no pushback. We do. What we’re saying is that our sense of the room, after presenting the relevant data and studies (covering everything from hiring decisions to the delivery of healthcare services) is that people want to hear us out, learn more. It is this moment—when people are sitting up, less likely to be defensive and default to “I get it,” “I’m rational,” or “I’m colorblind”—that our presentation takes a structural turn. Focusing specifically on law firms and the hiring context, we demonstrate how we do so.

3. The Structural Turn

A good way to setup the structural turn is to guide the audience into realizing the limitations of focusing on individual-
level-only strategies. For instance, one could trial-balloon the possibility of using implicit bias tests (like the IAT) as a screening device, a kind of anti-racist polygraph test, to prevent lawyers with high implicit bias scores from joining the firm. But the near universal reaction to such a countermeasure is deep unease. We then validate that reaction by making clear that nearly all serious scientists reject such a use of the IAT, and we do as well.

As another example, we might ask out loud whether receiving education, in the form of some entertaining implicit bias lecture (with CLE included), will suffice to defend against implicit bias. Again, the near universal reaction is deep skepticism. People know that listening to a talk, even delivered by experts, will not lead to practical changes in behavior necessary to drive change. Here, too, we validate that reaction and confirm with the maxim that “good intentions are not good enough.”

Part of what we are attempting to do by getting to “no” with respect to the preceding two inquiries is getting to “yes” with respect to a discussion about structures. What’s striking to us is that, with the preceding guiding questions about the limitations of individual-level interventions, law firms are quick to take the structural turn themselves, without any didactic urging on our part. Which is to say, in the context answering the prior questions in the negative (“No, let’s not use IAT as a polygraph” and “No, implicit bias education is insufficient to defend against implicit bias”), people invariably say things like: “We need to look at how we interview people,” “We need to examine our hiring criteria,” “We need to think about how and where we look for talent.” In other words, participants in the session themselves initiate the structural turn by inviting a discussion about policies, procedures, and practices.

To organize that discussion, we encourage law firms to view that hiring processes are machines that they construct, command, and control using various algorithms. As you will see, what we’re trying to do is to get firms to see the role they play—structurally—creating or not creating conditions of possibility for the diversity they say they. Often, we start the conversation with Figure []).
We encourage the audience to look at the image. We tell them that the image depicts a hiring machine and that the function of this machine is to produce “pools” from which the firm will hire associates. We then draw attention to the machine various inputs, Input A, Input B, Input C, and Input D. We next emphasize that firms have to specify and calibrate these inputs and that how they do so will determine the kinds of pools the machine produce.

We then turn to Figure [], which depicts two pools, Pool 1 and Pool 2.
Imagine that you’re a law firm partner and we asked you: At which one of these pools would you like to fish? You might begin by fighting the hypothetical and asking for more details. But supposed we responded by saying that, like with many law school puzzles, assume are to assume that you have imperfect information. What you know is that, demographically speaking, the pools look the way they look. Once more we ask: At which one of the pools do you want to fish—Pool 1 or Pool 2? Our surmise is that, like the people we have engaged in this way, you would pick Pool 2.

Our next move is to link these pools to our hiring machine along the lines Figure[] does.

![Hiring Machine Diagram](image)

By this time, everyone already appreciates that Pool 1 looks very different than Pool 2. Whereas the latter is racially diverse, the former is not. Everyone also appreciates that if a firm is looking to add Black fish to the ranks, there’s only two of them in Pool 1. Finally, everyone gets—and we explicitly state that—precisely because they are only two Black fish in Pool 1, not every firm can have one. Thus, most firms will return home professing a commitment to diversity but frustrated that they were unable to realize it because of the lack of diversity in Pool 1.

This is the point in which we remind our audience that firms don’t just happen to find themselves at Pool 1. Firms put themselves there by a structural decision they make about how to
construct and calibrate their hiring machine—in other words, by what they use as (and how they calibrate) Input A, Input B, Input C, and Input D. As Figure [] indicates, hiring machines can be constructed and calibrated to produce very different kinds of pools.

We are careful to note that there are real forces external to firm governance (for example, structural racism in society generally and in law school admissions policies, curriculum, and institutional cultures) that shape the “pools” from which law firms might hire. Our point is that as we recognize those external dynamics, we should also recognize that hiring pools aren’t fully exogenous to firm decision-making; they are also endogenous to firm decision-making as a function of how firms specify and calibrate Input A, Input B, Input C, and Input D of their hiring machine.

To make the preceding points inputs and pools more legible, we sometimes invoke data revealing the racial makeup of federal clerks. The table below captures the basic story.

<table>
<thead>
<tr>
<th>Racial/Ethnic Diversity Among Federal Clerks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nati</td>
</tr>
</tbody>
</table>

All figures exclude graduates for whom race or ethnicity was not reported. The Asian/Pacific Islander category includes East Indian/Pakistani and Native Hawaiian graduates. The option of reporting more than one race/ethnicity was introduced in 2001.

Source: National Association for Law Placement, Inc. (NALP)

Even the most cursory look at the data illustrates that the federal clerkship pool is not particularly diverse. We stress that when law firms rely heavily on clerkship as a job criterion, they are making a decision to fish in a pool that looks like Pool 1 from the Figure [].

At this point, we have primed the pump for firms to self-interrogate with respect to how they are specifying their inputs. We ask them to imagine that we are building up Input A, just one of the inputs that does into the firm’s hiring machine. Since every law firm will have “Merit Module,” a conception of merit
that drives its hiring, we tell to stipulate that Input A is the firm’s Merit Module. We then ask them to identify the top four criteria firms employ to hire associates. Typically, the answers are: grades, law review membership, writing sample, and law school rank. To stay with the machine metaphor, we encourage to think of Merit Module as its own machine (inside the Hiring Machine) that the firm also builds, controls, and commands. That machine, Figure [], looks like this:

![Diagram of a machine with inputs: Grades, Law Review, Writing Sample, Law School Rank.]

The analysis we typically perform upon showing this image is precisely the same as the analysis we performed with respect to the larger Hiring Machine within which the Merit Module is situated. Which is to say, we begin by noting that none of the criteria that comprise the Merit Module is preordained and all of them are part of a larger structural apparatus, the firm’s pool-producing Hiring Machine. Because firms exercise institutional agency about whether and how to employ the four criteria, they are also—self-consciously or not—exercise agency about the kinds of pools they want to create. This insight is important to highlight not only because these and other merit criteria often embed implicit biases and otherwise produce racial disparities, but also because merit criteria are not always useful proxies for the requirements of the job. *Washington v.*
Davis, the case we discussed earlier, is a useful case for illustrating these pitfalls.

As you might remember, a profound problem in Washington v. Davis was the fact that the test the Washington D.C. Police Department employed to hire its police officers (Test 21) was having a disparate impact on African American applicants. The Court in that case refused to require the police department to “validate” the test in the sense of requiring the department to demonstrate that Test 21 predicted on-the-job performance. Had the Court imposed that requirement, the government would likely have had to jettison the test.

We use the Davis example to encourage firms to think about whether their hiring criteria are sufficiently validated to function as on/off screening device switches. We ask them to think about whether their hiring criteria would pass something like a “strict scrutiny” review in the sense of there being a “compelling justification” for their utilization of a particular criterion and that the criterion is “narrowly tailored” to on-the-job requirements.

Note, again, that our ask of law firms here, like our ask of themselves with respect to our discussion about where to fish, is a structural one. It is not about changing people’s individual bias scores. We are encouraging firms to see that, at the level of organizational governance, they play a critical role shaping the diversity they want. To return to the metaphor of hiring pools, those pools are not just “out there” as fully formed configurations from which law firms hire talent. Firms help to produce the very pools in which they fish.96

C. LAW: REFORMING DOCTRINE

We now turn to a final context to describe briefly how taking implicit bias seriously can have significant structural

96 We recognize that there’s plenty with which one might quarrel about what we have said about race and law firm. And certainly, one can query whether our approach is sufficiently radical or progressive at all. But our investment here is less to defend any specific claim we have made about race and law firms and more to provide a domain-specific example of how a discussion that begins with the mind sciences—and specifically implicit biases—can culminate in an engagement that focuses on the structural dimensions of law firm governance. From that vantage point, our hope is that we’ve persuaded you that implicit bias education can be one vehicle (not the only or necessarily the most important vehicle) through which to teach the firm to think structurally about race in ways that have practical implications for law firm hiring.
implications in reforming legal doctrine. Although a comprehensive review is beyond the scope of this essay, we offer two examples, both of which reflect structural problems of race, capital punishment and Batson challenges.

In *McCleskey v. Kemp* (1987), the United States Supreme Court was presented with substantial statistical evidence showing racial disparities in who was sentenced to death. Nevertheless, the Court declined to find an Eighth Amendment violation on statistics alone. It explained:

At most, the [statistical] study indicates a discrepancy that appears to correlate with race. Apparent disparities in sentencing are an inevitable part of our criminal justice system. . . . Where the discretion that is fundamental to our criminal process is involved, we decline to assume that what is unexplained is invidious.

[W]e hold that the [statistical] study does not demonstrate a constitutionally significant risk of racial bias.

But recently, state courts have approached the similar question differently, under state constitutional law, in part because of implicit bias. Take, for example, the case of *Washington v. Gregory* (2018), in which the Washington Supreme Court took seriously its responsibility to reform law “in light of ‘advances in the scientific literature.’” The court then took “judicial notice of implicit and overt racial bias against black defendants in this state,” in order to reach the conclusion that “the association between race and the death penalty is not attributed to random chance.” The court explained, “We need not go on a fishing expedition to find evidence external to [the statistical] study as a means of validating the results. Our case

---


*481 U.S. 279 (1987).*

*Id. at 312–313.*

*427 P.3d 621 (Wash. 2018).*

*Consistent with a school of thought called behavioral realism, the court explained: “where new, objective information is presented for consideration, we must account for it. Therefore, Gregory’s constitutional claim must be examined in light of the newly available evidence presented before us.”*
law and history of racial discrimination provide ample support.”\textsuperscript{102}

Back in 1987, statistical disparities alone were not enough to persuade the US Supreme Court of a constitutional violation. But by 2018, a state Supreme Court could see the same statistical disparities, recognize implicit bias as one of the mechanisms by which racial bias could help to manifest those disparities, and use that revised understanding to justify a turning away from the conventional wisdom of \textit{McCleskey}.\textsuperscript{103}

Here’s another example that features the work of the state of Washington. Ever since the Supreme Court’s decision in \textit{Batson v. Kentucky} (1986), the deliberate striking of jurors on the basis of their race has been constitutionally prohibited. But proving up that purposeful state of mind – which evidenced a privileging of the “perpetrators perspective” -- proved to be extremely difficult given how easy it was for prosecutors and attorneys to provide plausible reasons for exercising peremptory challenges.

In 2018, the Washington Supreme Court decided that proving up this purposeful discrimination would no longer be necessary to state a constitutional objection. As codified in their General Rule 37,\textsuperscript{104} the question has now become whether “an objective observer could view race or ethnicity as a factor in the use of the peremptory challenge.”\textsuperscript{105} To see the role that implicit bias might have played in adopting this approach, notice how the state Supreme Court elaborated on what this “objective observer” should be aware of. General Rule 37(f) specifically states: “[A]n objective observer is aware that \textit{implicit, institutional, and unconscious biases}, in addition to purposeful discrimination, have resulted in the unfair exclusion of potential jurors in Washington State.”\textsuperscript{106}

\begin{notes}
\item[102] Id. at 635.
\item[105] Ibid., 470. See also Wash. St. Ct. Gen. R. 37(e).
\item[106] Wash. St. Ct. Gen. R. 37(f) [emphasis added].
\end{notes}
What’s fascinating is that other states, including California,\textsuperscript{107} Connecticut,\textsuperscript{108} and New Jersey,\textsuperscript{109} have followed Washington’s lead. They all move away from an exclusive focus on the subjective intent of the prosecutor and move toward an objective interpretation. Moreover, they all emphasize the importance of implicit bias as part of what a reasonable, objective observer would understand. And recall our discussion of Charles Lawrence’s seminal call for a “cultural meaning” test, which also relied on an objective observer. All of this is to demonstrate that taking implicit bias seriously has been at least consistent with, if not a useful driver of, legal reform that has embraced a structural approach.

**CONCLUSION**

At bottom, this Essay argues that the Left should take implicit bias seriously for its potential role in structural transformation. As a matter of theory, implicit bias can articulate, provide additional evidence for, and support crucial concepts and Critical Race Theory. For instance, it can help to

\textsuperscript{107} See AB 3070, which requires bars “the use of group stereotypes and discrimination, whether based on conscious or unconscious bias, in the exercise of peremptory challenges.” Assem. Bill 3070, ch. 318 (Cal. 2020), codified at Cal. Civ. Proc. Code § 231.7, Sec. 1(c) (emphasis added). The law asks whether “there is a substantial likelihood that an objectively reasonable person would view race [and other protected categories] as a factor in the use of the peremptory challenge.” Cal. Civ. Proc. Code § 231.7, Sec. 2(d) (emphasis added).

\textsuperscript{108} State courts must now ask whether the peremptory challenge “as reasonably viewed by an objective observer, legitimately raises the appearance that the prospective juror’s race or ethnicity was a factor.” Sec 5–12(d), in Official 2023 Connecticut Practice Book (Revision of 1998): Containing Rules of Professional Conduct, Code of Judicial Conduct, Rules for the Superior Court, Rules of Appellate Procedure, Appendix of Forms, Notice Regarding Official Judicial Branch Forms, Appendix of Section 1–9B Changes (Hartford: The Commission on Official Legal Publications, 2023), 180. Similar to the State of Washington’s approach, the objective observer “is aware that purposeful discrimination, and implicit, institutional, and unconscious biases, have historically resulted in the unfair exclusion of potential jurors.” “Sec. 5–12(e)” (emphasis added).

\textsuperscript{109} the State of New Jersey to no longer require a showing of “purposeful discrimination.” Instead, courts must now ask whether “a reasonable, fully informed person would view the contested peremptory challenge” to be based on a protected social category. N.J. Ct. R. 1:8–3A (emphasis added). The Official Comment lists reasons that are presumptively invalid because they are historically associated with “improper discrimination, explicit bias, and implicit bias.” Id., Official Comment (3) (emphasis added).
illustrate not only how race is socially constructed, but also the relationship between the social construction of race and the material conditions of racial inequality. More specifically, implicit bias education can put into sharp relief the relationship between social meanings of race and structural racism. One way to understand that relationship would be to say that negative ideas about racial groups (ex ante) lays the groundwork for and helps to legitimize the racial subordination of those groups (ex post).

As a matter of practice, implicit bias makes it more difficult for people to relegate racism to the domain of ancient history. Implicit bias evidence suggests that in the “here and now” negative social meanings about race persist (for example, the perception of people of Asian descent as inassimilable foreigners) and help explain various forms of racism (for example, anti-Asian violence people experience today). As applied to redesigning the law firm, this orientation strengthens moral ownership of the problems of implicit bias and encourages law firms to experiment with structural changes to how they operate.

Finally, as a matter of law, taking implicit bias seriously has already helped courts pipit away from the perpetrator perspective and toward more structural understandings of various racial problems. The movements by state court in the fields of capital punishment and peremptory challenges demonstrate beautifully how an approach that takes the mind sciences seriously, including implicit bias, can produce legal reform with deeply structural sensibilities.

In providing this robust defense of implicit bias’s possibility, we reiterate that this is possibility, not guarantee. We recognize that in various domains, the turn to implicit bias may ultimately fail to persuade or may produce only modest results.