Implicit Bias, Behavioral Realism, and the Purposeful Intent Doctrine

Jerry Kang*

Version 0.4 | May 18, 2022

forthcoming in Oxford Handbook of Race and the Law (Devon Carbado ed. 2023).

Thanks to: Anthony G. Greenwald, Calvin Lai
Paper given at WashU Law.

Research Assistance (Angel Lee).

* Distinguished Professor of Law and (by courtesy) Asian American Studies, UCLA. <kang@law.ucla.edu> <http://jerrykang.net>.

INTRODUCTION

Implicit social cognition examines the mental processes that affect social judgments without full self-awareness or control. Over the past quarter century, scientific findings in implicit social cognition generally and implicit bias specifically have challenged our traditional understandings of racial discrimination.[1] Legal analysts and mind scientists have argued that these new empirical findings should influence the development of anti-discrimination law. This chapter provides a brief primer on implicit bias and examines how its discovery has catalyzed legal reform through a school of thought called "behavioral realism."

I. SCIENCE OF IMPLICIT BIAS

A. The Idea

The *idea* of "implicit bias" can be unpacked by examining each word that makes up the term.[2] The noun "bias" just means deviation from some baseline of comparison that is accepted as neutral or appropriate. In social psychology, bias toward a category is typically understood in terms of attitudes or stereotypes. An "attitude" is an overall evaluative valence toward a category, which ranges from positive to negative. To take an uncontroversial example from the animal world, some people prefer dogs to cats. Their attitude toward dogs is positive whereas their attitude toward cats is less so and perhaps even negative. By contrast, a "stereotype" is an attribute or trait that is probabilistically associated with a category. Consider how the trait "loyal" might be more associated with dogs than cats, at least for dog-lovers. As another example, consider how the stereotype of "vicious" might be associated more with pit-bulls than retrievers.

The adjective "implicit" is best understood in contrast to "explicit." "Explicit" need not be graphic or extreme as that word is commonly used in terms such as "explicit lyrics." Instead, it's better to understand "explicit" as being

[1] *See generally* Kristin A. Lane, Jerry Kang & Mahzarin R. Banaji, Implicit Social Cognition and Law, 3 ANN. REV. L. & SOC. SCI. 427 (2007).

[2] Much of Part I draws on Jerry Kang, *What Judges Can Do about Implicit Bias*, 57 Ct. Rev. 78, 78-81(2022).

subject to direct introspection. Let's return to household pets. Suppose I ask you what you think about dogs. Suppose further that you love dogs and have a puppy nuzzling in your lap right now. The fact that you're able to ask yourself the question and get a clear, immediate answer back through direct introspection means that you have accessed and reported an explicit attitude. And because there isn't much stigma about loving dogs in American culture, there's little social pressure for you to hide that explicit attitude from others or from yourself.

By contrast, an "implicit" bias is an attitude or stereotype that is *not* readily subject to direct introspection. In other words, we cannot easily[3] measure implicit social cognitions simply by asking ourselves direct questions about our own social cognitions. Accordingly, direct introspection can make us only dimly self-aware of these cognitions and how they might influence our judgments and behavior.

B. The Instrument

Because explicit bias is subject to direct introspection, it is typically measured through a survey instrument, with the hope that participants answer honestly. By contrast, because implicit bias is not readily subject to direct introspection, it must be measured some other way. Experimental social psychologists have developed over eighteen instruments, as usefully organized in a recent inventory.[4] The lengthy list immediately encourages us not to conflate the *idea* of implicit bias from any particular *instrument* by which it is measured, such as the well-known Implicit Association Test (IAT).

---

[3] Some recent research suggests that implicit social cognitions may be preconscious, subject to some forms of introspection when it is guided by concrete stimuli and directions to pay attention to immediate affective responses. *See generally* Adam Hahn & Alexandra Goedderz, *Trait-Unconsciousness, State-Unconsciousness, Pre-Consciousness, and Social Miscalibration in the Context of the Implicit Evaluation*, 38 Social Cognition S115 (2020) (supplement).

[4] *See* Anthony G. Greenwald & Calvin K. Lai, *Annual Review of Psychology: Implicit Social Cognition*, 71 Ann. Rev. of Psychology 419, 423-24 (2020) (providing three categories: (1) the Implicit Association Test (IAT) and its variants; (2) priming tasks (where brief exposure to priming stimuli facilitates or inhibits subsequent reactions); and (3) miscellaneous other tasks including linguistic or writing exercises); Calvin K. Lai & Megan E. Wilson, *Measuring Implicit Intergroup Biases*, 15 Social & Personality Psychology Compass 1 (2021).

The IAT is the most used and best validated instrument for measuring implicit bias. It's essentially a videogame requiring fast sorting of stimuli flashed on a computer screen. These stimuli represent either two social categories (e.g., White people versus Black people represented by photos) or two social cognitions (e.g., positive versus negative attitude represented by words). The stimuli are positioned on the screen in a way that demand keyboard responses that are either consistent or inconsistent with our implicit social cognitions.

For instance, if we have a more positive implicit attitude toward White people, then we should respond more quickly when White faces (representing White people) and Good words (representing a positive attitude) are paired together on the same response key as compared to the opposite arrangement of Black faces and Good words. This faster response might be just a few hundred milliseconds. In the course of an IAT, these raw reaction time deltas are repeatedly measured through multiple blocks of trials, mathematically processed, and transformed into statistical units. These reaction time latencies roughly signal the comparative strength of the underlying implicit association. Since the IAT has been described extensively elsewhere,[5] I won't go into greater detail here. But if you're unfamiliar, try taking one for free, anonymously, at Project Implicit.[6]

Millions of people have done so since the late 1990s, in the United States and around the globe. The first systematic analysis of the pervasiveness and correlates of implicit attitudes and stereotypes, as measured by the IAT, was conducted by Brian Nosek and colleagues in 2007 (describing data collected on 17 different tests at Project Implicit during 2000-2006).[7] They found that implicit bias—as measured by the IAT at Project Implicit—was pervasive. I have it. You have it. Not in precisely the same amounts, toward the same categories, but we all have it. Implicit bias was also found to be generally larger in magnitude than explicit

[5] *See, e.g.*, Kristin A. Lane, Jerry Kang, and Mahzarin R. Banaji, *Implicit Social Cognition and the Law*, 3 ANNU. REV. LAW SOC. SCI. 19.1, .2-.3 (2007); Jerry Kang & Kristin Lane, *Seeing through Colorblindness: Implicit Bias and the Law*, 58 UCLA L. REV. 465, 472-73 (2010).

[6] *See* <http://projectimplicit.org>.

[7] Brian A. Nosek et al., *Pervasiveness and Correlates of Implicit Attitudes and Stereotypes*, 18 EUR. REV. SOC. PSYCHOL. 1 (2007).

bias self-reported on surveys. At least in reaction times, we are not as neutral or "blind" as we claim to be.

Recently, Kate Ratliff and colleagues updated the Nosek analysis with Project Implicit data from 2007-2015. They again found that implicit bias "favoring culturally dominant or societally valued groups" remains pervasive and stronger in magnitude than explicit bias. They also found that "ingroups are evaluated more positively than outgroups."[8] This finding reminds us to be mindful of not only outgroup derogation but also ingroup favoritism, an important source of modern discrimination.[9] Overall, these large dataset analyses are consistent with IAT data generated from experiments conducted in hundreds of laboratories around the world over the past quarter century.

To sum up: When asked if we are colorblind, we may sincerely answer that we judge people only by the content of their character not the color of their skin. In other words, race doesn't matter. But at least on the IAT sorting game, it isn't so. We generally reveal an implicit attitude in favor of Whites over Blacks. We reveal an implicit stereotype that associates weapons more with Blacks than Whites. We implicitly associate "American" more with Whites than East Asians. And so on. People of color, by virtue of their social category, are not magically immune. For example, approximately one-third of Black Americans themselves reveal an implicit attitude in favor of Whites over Blacks. We may wish it otherwise, but wishing doesn't make it so.

C. The Impact

The fact that implicit bias measured by some instrument is *statistically* significant does not necessarily mean that it is *socially* significant, worthy of individual or institutional response. For that to be the case, we'd want evidence

[8] *See* Kate A. Ratliff et al, *Documenting Bias from 2007-2015: Pervasiveness and Correlates of Implicit Attitudes and Stereotypes II* (unpublished pre-print) at 2. The meta-analytic effect size for implicit bias was Cohen's d =.80 (by convention called "large") as compared to explicit bias of d = .51 (by convention called "medium"). *Id.* at 21.

[9] *See* Anthony G. Greenwald & Thomas F. Pettigrew, *With Malice toward None and Charity for Some: Ingroup Favoritism Enables Discrimination*, 69 American Psychologist 669 (2014). Even if every ingroup favors itself equally, the population and resource advantage of certain groups will lock in a net advantage indefinitely.

that implicit bias causes or at least correlates with real-world discriminatory behavior, by which I mean worse treatment on the basis of a social category (e.g., Black) as compared to the baseline category (e.g., White). For the IAT, which is the most widely-used instrument, multiple meta-analyses demonstrate that implicit bias correlates with discriminatory behavior—but to a small degree.

A "meta-analysis" is an analysis of analyses. It takes all studies that can be found in the relevant domain and synthesizes their findings into a single composite number, usually the "effect size." In implicit social cognition, the effect size is measured by the degree of correlation between an implicit bias measure and discriminatory behavior. The correlation is a measure of a linear relationship between two variables, such as age and height, height and weight, or in this case implicit bias and discriminatory behavior. It is quantified by Pearson's r, which runs from 0, which means no relationship between bias and behavior, to ±1, which means a perfectly linear positive or negative relationship.

All meta-analyses show that IAT scores predict intergroup discriminatory behavior at a low level. The range of r values goes from 0.10 to 0.24,[10] which by disciplinary convention is called a "small-to-moderate" effect size.[11] The modest effect sizes found should not be surprising given how difficult it is to measure accurately the strength of mental associations through reaction times and how difficult it is to measure accurately whether some behavior or

[10] *See* Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity*, 97 J. PERSONALITY & SOC. PSYCHOL. 17, 19–20 (2009) (r=.024 for Black/White bias); Frederick Oswald et al., *Predicting ethnic and racial discrimination: A meta-analysis of IAT research*, 105 J. Personality & Soc. Psychol. 171 (2013) (r=.15 on Black/White implicit bias); Benedikt Kurdi et al., *Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis*, 74 American Psychologist 569 (2019) (personal communication to Anthony Greenwald that r = .10 for behavior in the *combined* domains of Black/White, gender, sexual orientation, weight, and disabilities). The highest estimate (r=0.24) would mean that an IAT score predicts approximately 5.6% of the statistical variance—a formal measure of the spread of a distribution—in the discriminatory behavior variable.

[11] By convention, r = .10 is called "small," and r = .30 is called "moderate."

judgment is discriminatory, especially in realistic situations. Noisy measures of any two variables—in this case, bias and behavior—necessarily weaken the strength of the measured relationship that can be found between them.

So, how should we understand the social significance of small effect sizes? One pragmatic approach is to ask, "As compared to what?" First, compare implicit bias scores to explicit bias scores. Implicit bias predicts intergroup discriminatory behavior *better* than explicit measures of bias.[12] In other words, IAT measures are more predictive than the answers that people give to explicit questions about bias. So, if we conclude that implicit bias scores tell us nothing useful, should we abandon asking potential jurors about explicit bias in *voir dire*?

Second, compare the effect sizes of implicit bias scores with the effect sizes of medical commonsense. Someone in your family may have taken aspirin to reduce the chances of death by heart attack. You may have told your teenager (or been told by your parents) never to smoke because of the risks of lung cancer. You may have shielded your child from low-level lead exposure for fear of its impact on childhood IQ. But all of these commonsense behaviors turn out to be based on correlations lower than the correlation found between implicit bias and discriminatory behavior.[13] So, if we confidently announce that small effect sizes can be ignored, are we going to stop worrying about exposing children to lead? Stated more sharply, are we going to stop worrying about lead exposure for *our own* children? And, if the answer is no, on what grounds do we worry about *our own* children's IQ but ignore implicit bias-actuated discrimination suffered by *others*?

Finally, as has been formally modeled, small differences in treatment and evaluation can accumulate in path-dependent ways over long periods of time and across large populations to produce surprisingly significant social dis-

[12] *See*, *e.g.*, Greenwald et al., *supra* note 13, at 73 tbl. 3 (finding that implicit attitude scores predicted behavior in the Black/White domain at an average correlation of r=0.24, whereas explicit attitude scores had correlations of average r=0.12); Kurdi et al, *supra* note 13 (finding that implicit biases provide a unique contribution to predicting behavior (β = .14) and does so more than explicit measures (β = .11)).

[13] *See* Gregory J. Meyer et al., *Psychological Testing and Psychological Assessment: a Review of Evidence and Issues*, 56 American Psychologist 128, 130 tbl. 1 (2001).

parities.[14] Little things accumulate. In sum, implicit biases are attitudes and stereotypes whose existence and influence we are only dimly aware of through direct introspection. They correlate in small ways with subtle discriminatory judgments and behavior. Nevertheless, these small impacts can produce large consequences over the timeline of an individual's life and across large groups.[15]

How should this more sophisticated, upgraded understanding influence the law?

II. LEGAL UPTAKE

A. Behavioral Realism

One approach is to embrace "behavioral realism," a school of thought that combines the traditions of legal realism and behavioral science. The first public call for behavioral realism appeared in the Harvard Law Review in 2005.[16] It was explored in depth in a collection of articles published in the California Law Review in 2006, in which pairs of social psychologists and legal scholars authored papers to consider the legal and social significance of new findings from social psychology.[17] Stated succinctly, behavioral realism insists that law should incorporate a scientifically up-to-date model of human behavior.

In prior work, I've elaborated a simple three-step process of what this entails.[18] First, be on the lookout for a

[14] *See* Anthony G. Greenwald et al., *Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects*, 108 J. Person. & Soc. Psych. 553 (2014); Jerry Kang et al., *Implicit Bias in the Courtroom*, 59 UCLA L. Rev. 1124, 1143, 1151-52 (2012); Jerry Kang & Mahzarin Banaji, *Fair Measures: A Behavioral Realist Revision of 'Affirmative Action,'* 94 CALIF. L. REV. 1063, 1073 (2006).

[15] To be clear, implicit biases are not the only biases that matter to basic fairness. Explicit biases persist. Systemic biases remain and perpetuate inequalities established long ago that continue to structure opportunity going forward. For more discussion, see Jerry Kang, Why Implicit Bias? (draft).

[16] *See* Jerry Kang, *Trojan Horses of Race*, 118 Harv. L. Rev. 1489, 1494 n.21 (comparing and contrasting behavioral realism with rational choice theory and behavioral law and economics). *See id.* at 518 (arguing why critical race theorists should embrace behavioral realism).

[17] *See* 94 Calif. L. Rev. 945-1190 (2006).

[18] *See, e.g.*, Jerry Kang, *Rethinking Intent and Impact: Some Behavioral Realism about Equal Protection*, 66 Alabama L. Rev. 627-51 (2015);

more accurate, upgraded model of human behavior that comes from the sciences, including implicit social cognition. Second, compare that upgraded model to the "commonsense," legacy understanding embedded within the status quo. Even when no such legacy model is apparent, careful excavation will generally reveal background assumptions so obvious or uncontroversial as to be nearly invisible. Third, when the gap between the upgraded and legacy models grows too large (however defined), revise the law to take account of the upgraded model. If that can't be done, then lawmakers should explain transparently their good reasons why.[19]

Notice how this three-step process does not rely on any particular theory of corrective or distributive justice even as applied to anti-discrimination law. Instead, it trades on minimalist normative commitments—(1) instrumental rationality (by revising the law to function under a more accurate model of human behavior), and (2) anti-hypocrisy (by prompting law's re-design when law's publicly announced objectives cannot be achieved because of reliance on an outdated understanding). Notwithstanding the simplicity of this approach, behavioral realism has had a significant impact on the law. In particular, it has weakened antidiscrimination law's preoccupation with the "intent doctrine."

B. The (Purposeful) Intent Doctrine

Every constitutional law student learns of the "intent doctrine" through *Washington v. Davis* (1976).[20] In that case, the question presented was whether a particular qualifying test that produced a disparate impact on Black police officer candidates violated their federal constitutional equal protection rights. The Court explained that because there was no purposeful intent to harm Black candidates, there was no constitutional infirmity. This interpretation was strengthened three years later in *Personal Administrator v.*

Jerry Kang, *The Missing Quadrants of Anti-discrimination: Going Beyond the "Prejudice Polygraph,"* 68 J. Social Issues 314-27 (2012); Jerry Kang & Kristin Lane, *Seeing through Colorblindness: Implicit Bias and the Law*, 58 UCLA L. Rev. 465–520 (2010); Kristin A. Lane et al., Implicit Social Cognition and Law, 3 ANN. REV. LAW & SOC. SCI. 427, 440 (2007).

[19] Many details have been omitted. For greater elaboration, see Lane & Kang, supra note _.

[20] 426 U.S. 229 (1976).

Feeney (1979).[21] This is commonly called the "intent doctrine," which is unfortunately named given the vagueness of the word "intent." After all, every student of the Model Penal Code learns of at least four distinct levels of culpability in substantive criminal law: purposely, knowingly, recklessly, and negligently.[22] What's critical to understand is that for federal constitutional equal protection, only a finding of *purposeful* intent matters. Knowing, reckless, or negligent state action is incognizable.[23]

Interestingly, the purposeful intent requirement is not required in all civil rights laws. The best-known counterexample comes from employment discrimination prohibited by federal law under Title VII of the 1964 Civil Rights Act. By its own terms, that statute prevents adverse employment actions "because of" a protected social category, such as race.[24] As such, Title VII unambiguously prohibits disparate treatment of the kind that would evince a purposeful intent to harm racial minorities under *Washington v. Davis*.

But the statute has been read to encompass far more. In *Griggs v. Duke Power* (1971), the Supreme Court adopted a "disparate impact" theory of liability that proscribed facially neutral employment policies or practices that produced a disparate impact on racial minorities unless some

[21] For thoughtful and historically inflected analyses of these cases, see Ian Haney-Lopez, *Intentional Blindness*, 87 NYU L. REV. 1779 (2012); Reva Siegel, *Foreword: Equality Divided*, 127 HARV. L. REV. 1 (2013). Siegel sees *Feeney* as a decisive move to "restrict the ways that evidence of foreseeable impact could be used to prove unconstitutional purpose." *Id.* at 17.

[22] *See* Model Penal Code § 2.02 (General Requirements of Culpability). *Purposely* mean that it is a person's "conscious object to engage in conduct of that nature or to cause such a result"); *knowingly* means a person "*is aware* that his conduct is of that nature or that such circumstances exist" or is "*practically certain* that his conduct will cause such a result"; *recklessly* means that a person "*consciously disregards* a substantial and unjustifiable risk" that "involves a gross deviation from the standard of care that a reasonable person would observe in the actor's situation"; negligently means a person "*should be aware* of a substantial and unjustifiable risk" that "involves a gross deviation from the standard of care that a reasonable person would observe in the actor's situation." *Id.* (emphasis added).

[23] This is consistent with how the MPC uses "intentionally" or "with intent". *See* MPC § 1.13(12) (Definitions) ("'intentionally' or 'with intent' means purposely)

[24] *See* 42 U.S.C. § 2000e-2(a).

business necessity could be demonstrated.[25] This theory of liability was later ratified by Congress in 1991.[26]

The *Griggs* holding stands as an early existence proof that civil rights laws don't have to be limited to purposeful intent. It further demonstrates that ambiguous statutory text, such as Title VII's "because of," can be interpreted to do just that. But *Griggs* remains more the exception than the rule in antidiscrimination law. Although decided five years before *Washington v. Davis*, the Supreme Court declined to follow the *Griggs* path for equal protection law. Purposeful intent remains the constitutional touchstone and arguably the initial presumption in interpreting all civil rights laws.

Unfortunately, purposeful intent is both difficult to prove and ignores equality harms produced at lower levels of intent. That's why Critical Race Theorists have criticized the purposeful intent doctrine for privileging the "perpetrator perspective."[27] And civil rights organizations, such as the Equal Justice Society, have adopted as its mission "directly challenging the [Purposeful] Intent Doctrine."[28] All this tees up the important question: has the science of implicit bias, by way of behavioral realism, weakened our fixation on purposeful intent?

C. Chipping Away at (Purposeful) Intent

1. Federal Housing Discrimination

In the 2015 case of *Texas Department of Housing & Community Affairs v. Inclusive Communities Project, Inc.*,[29] the Supreme Court addressed whether the Fair Housing Act (FHA) countenanced a disparate *impact* theory of liability. The Supreme Court went through standard statutory interpretive moves, such as drawing on analogies with Title VII's disparate impact theory adopted in *Griggs v. Duke Power*. But at a

[25] *See* Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971).

[26] Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071, 1074 (codified at 42 U.S.C. § 2000e-2(k) (2006)).

[27] *See* Alan David Freeman, *Legitimizing Racial Discrimination Through Antidiscrimination Law: A Critical Review of Supreme Court Doctrine*, 62 Minn. L. Rev. 1049 (1978).

[28] *See* <https://equaljusticesociety.org/aboutus/mission/> (last visited March 23, 2022).

[29] 576 U.S. 519 (2015).

key moment of its reasoning, the Court (per Justice Anthony Kennedy) explained:

> Recognition of disparate impact liability under the FHA also plays a role in uncovering discriminatory intent. It permits plaintiffs to counteract *unconscious prejudices* and disguised animus *that escape easy classification* as disparate treatment. In this way disparate impact liability may prevent segregated housing patterns that might otherwise result from covert and illicit stereotyping.[30]

This explanation suggests that the Justices in the five-person majority had upgraded its model of discrimination to recognize the existence of "unconscious prejudices." In addition, they saw the theory of disparate impact—which could invalidate facially neutral housing policies or practices that nevertheless produce a disparate impact on racial minorities without any showing of purposeful intent to harm them—as a tool with which to respond to this more complex understanding. Partly because of this upgraded model of human behavior and decision-making, the Court felt more comfortable interpreting the ambiguous statutory phrase "because of"[31] within the FHA to include disparate impact liability. And this approach was taken in the face of strident dissents, including one from Justice Clarence Thomas, who argued that *Griggs* was wrongly decided and should not be allowed to metastasize to the law of fair housing.

This case is one prominent example in which behavioral realism about implicit bias facilitated movement away from the purposeful intent doctrine by *recognizing a disparate impact theory of liability* for the federal Fair Housing Act.

2. Federal Employment Discrimination

In the 2010 case *Kimble v. Wisconsin Dept. of Workforce Development*,[32] plaintiff raised a disparate *treatment* discrimination claim under Title VII of the 1964 Civil Rights Act. Before the rise of implicit social cognition, it would have been natural to assume that disparate *treatment* requires an explicit, purposeful intent to treat racial minor-

[30] *Id.* at 540 (emphasis added).
[31] 42 U.S.C. § 3604(a), § 3605(a).
[32] 690 F.Supp.2d 765 (E.D. Wis. 2010).

ities worse, especially to distinguish disparate treatment from *Griggs*'s disparate *impact* theory of liability. After all, under the legacy model, this was our commonsense understanding of how racial discrimination occurred.

But the plain text of the statute nowhere includes the words "explicit" or "purposeful." Instead, as already noted, Title VII merely requires "because of"—which by its own terms suggests only a *causal* connection between the adverse employment action and the plaintiff's protected social categories. As such, Title VII is textually open to a disparate treatment theory that goes beyond purposeful discrimination.[33]

With this broader possibility in mind, the Eastern District of Wisconsin pivoted away from purposeful intent and instead embraced racial causation. It explained "[n]or must a trier of fact decide whether a decision-maker acted purposively…. Rather, in determining whether an employer engaged in disparate treatment, the critical inquiry is *whether its decision was affected* by the employee's membership in a protected class."[34] Applying this legal understanding to the facts in a bench trial, the court observed that "when the evaluation…is highly subjective, there is a risk that supervisors will make judgments based on stereotypes of which they *may or may not be entirely aware*."[35] It noted that because of the ordinary process of categorical thinking, a supervisor may use stereotypes "*whether or not* the supervisor is *fully aware* that this is so."[36]

This explanation reflects an upgraded model of discrimination, which the court attributed to various second-

[33] *See id. at* 768 ("Nor must a trier of fact decide whether a decision-maker acted purposively or based on stereotypical attitudes of which he or she was partially or entirely unaware.").

[34] *See Kimble v. Wiscon. Dep't of Workforce Development*, 690 F.Supp.2d 765, 768-69 (E.D. Wis. 2010)) (emphasis added). *See also Ricci v. DeStefano*, 557 U.S. 579, 580 (2009) (making clear that malevolent or hostile motivations are not required for a disparate treatment claim). *See also* Linda Hamilton Krieger & Susan T. Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 CALIF. L. REV. 997, 1053-54 (2006); Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161, 1226 (1995).

[35] Id. at 775-76.

[36] Id. at 776.

ary sources, including law review articles embracing behavioral realism.[37] And on this basis, the court held that the plaintiff had made his prima facie case, that the defendant was not a credible witness, and that ultimately a disparate *treatment* violation had in fact taken place.

This is one example in which behavioral realism about implicit bias facilitated recognizing a disparate treatment theory of liability under Title VII that *does not require purposeful intent and instead demands only racial causation*.

3. Federal Civil Procedure

In the 2017 case of *Woods v. City of Greensboro*,[38] plaintiffs alleged federal civil rights violations under 42 U.S.C. § 1981 (equal contracting rights) because the city denied them certain economic development loans.[39] As has been recently clarified by the Supreme Court, § 1981 requires a showing that race was the "but-for" cause of the denial.[40] ("But-for" causation requires a showing that the discrimination would not have taken place but for the plaintiff's race. It is harder to demonstrate than a mere "motivating factor.")

Defendant moved to dismiss for failure to state a claim upon which the law grants relief, under Fed. R. Civ. Pro. 12(b)(6). Historically, this motion to dismiss, which occurs before discovery, was difficult to grant because courts accepted as true all factual allegations made in the complaint and merely checked for legal sufficiency. But since *Bell Atlantic Corp. v. Twombly*[41] and *Ashcroft v. Iqbal*, it has become easier to grant the motion because courts were instructed to independently examine the "plausibility" of the

[37] *See id.* at 776 (citing articles appearing in the Behavioral Realism Symposium, *supra* note **Error! Bookmark not defined.**).

[38] 855 F.3d 639 (4th Cir. 2017).

[39] 42 U.S.C. § 1981(a) ("All persons within the jurisdiction of the United States shall have the same right in every State and Territory to make and enforce contracts...as is enjoyed by white citizens....").

[40] *See* Comcast Corp. v. Nat. Assn. African American-Owned Media, 140 S.Ct. 1009, 1014 (2020) ("[P]laintiff bears the burden of showing that race was a but-for cause of its injury"). *See also id.* at 1019. [SWAT: check that this was clear in that circuit in 2017]

[41] 550 U.S. 544 (2007).

allegations made. Applying this plausibility standard, the district court granted the motion to dismiss.[42]

Surprisingly, the Fourth Circuit Court of Appeals reversed on the basis of an upgraded model of discrimination. It started its analysis by noting that "many studies have shown that most people harbor *implicit biases* and even well-intentioned people *unknowingly act* on racist attitudes."[43] Showing psychological sophistication, the court pointed out that the same actor may discriminate differently depending on the context: "it is unlikely today that an actor would *explicitly discriminate* under all conditions; it is much more likely that, where discrimination occurs, it does so in the context of *more nuanced decisions* that can be explained based upon reasons *other than illicit bias*, which though *perhaps implicit, is no less intentional*."[44] Finally, the court warned that: "there is thus a real risk that legitimate discrimination claims, particularly claims based on more *subtle theories* of stereotyping or *implicit bias*, will be dismissed should a judge substitute his or her view of the likely reason for a particular action in place of the controlling plausibility standard."[45] For these reasons, the court of appeals reversed and allowed the case to proceed to discovery.

This is one example in which behavioral realism about implicit bias altered judicial common sense of what's plausible, making a discrimination claim more likely to survive a motion to dismiss under federal civil procedure.

4. State Criminal Law (Policing)

In the 2019 case *Kansas v. Gill*,[46] a Kansas appellate court had to interpret a newly enacted state statute that prohibited "racial or other biased-based policing,"[47] which was defined as "the unreasonable use of race...by a law enforcement officer in deciding to initiate an enforcement

[42] 556 U.S. 662 (2009). For discussions about how implicit bias could infect the "plausibility" standard, see Kang et al., *supra* note 14, at 1159-64.

[43] *Id.* at 641. [Which citation?]

[44] *Id.* at 651-52.

[45] *Id.* at 652.

[46] 56 Kan. App. 2d 1278 (2019).

[47] Kan. Stat. Ann. § 22-4609.

action."[48] In this case, a police officer had approached two Black men in a SUV because they were "staring hard" at him, which approach led to the officer's smelling marijuana and discovering drugs in the car. The trial court found that the police officer had used race unreasonably, and on appeal the majority of the court of appeals affirmed.

What's fascinating is how the majority responded to the dissent's incendiary framing: "Before we brand an officer of the law…a racist, there ought to be evidence supporting such a serious charge."[49] The majority calmly replied that "*no one here is branding [the officer] a racist*."[50] Instead, invoking an upgraded model of discrimination, it framed the question presented as one of racial causation, whether the officer "let racial bias—*conscious or unconscious—affect* his initiation of enforcement action."[51] It observed that the statute did not expressly require "exceptionally horrific or despicable race-based behavior"; to the contrary, the "[l]egislature recognized that racial bias is not always overt, it is often *subtle*" and that federal law enforcement leadership had publicly recognized that "all people have some form of *unconscious racial biases*."[52]

This is one intriguing example in which behavioral realism about implicit bias helped a state court reject a purposeful intent requirement in interpreting a state statute that proscribed law enforcement's unreasonable use of race.

5. State Jury Selection (Per-emptory challenges)

In *Batson v. Kentucky* (1986),[53] the Supreme Court held that purposeful discrimination to strike jurors on the basis of race violated federal equal protection guarantees. In order to ferret out such purposeful discrimination, the Court implemented a familiar burden shifting mechanism. First, the party raising the constitutional claim must establish a prima facie case. Second, the opposing party must provide a race-neutral explanation for the peremptory challenges. Third, and most important, the court must decide on the ba-

[48] Id. § 22-4606(d).
[49] *Id.* at 1288 (Powell, J., Dissenting).
[50] *Id.* at 1286 (emphasis added).
[51] *Id.* at 1286-87 (emphasis added).
[52] *Id.* at 1287 (emphasis added).
[53] 476 U.S. 79 (1986).

sis of all the evidence whether the challenging party showed "purposeful discrimination."[54]

Unfortunately, this three-step procedure rarely found what it was looking for. This should not be surprising given the ease with which lawyers can construct non-racial explanations, plausible to the court and to themselves, about why they struck a juror who just happened to be a racial minority. As explained by Justice Stephen Breyer in his concurrence in *Miller-El v. Dretke*,[55] "step three [of] *Batson* asks judges to engage in the awkward, sometime hopeless, task of second-guessing a prosecutor's instinctive judgment—the underlying basis for which may be *invisible even to the prosecutor*."[56]

Nearly four decades later, the Washington Supreme Court took advantage of the flexibility permitted in the local implementation of *Batson* and pivoted away from the standard procedure.[57] For the crucial third step, instead of demanding a factual subjective "purposeful discrimination," the court substituted a more interpretive objective reasonable person standard. It did so prospectively via judicial rulemaking and adopted General Rule 37,[58] and then retroactively, in *Washington v. Jefferson* (2018).[59] Specifically, Washington's version of step three of the *Batson* process now asks whether "an *objective observer* could view race or ethnicity *as a factor* in the use of the peremptory challenge."[60] What's especially fascinating is that this hypothetical objective observer benefits from a fully upgraded model of discrimination. General Rule 37(f) expressly states: "Nature of Observer. For purposes of this rule, an objective observer is aware that *implicit, institutional, and unconscious biases*, in addition to purposeful discrimi-

[54] *Id.* at 98.

[55] 545 U.S. 231 (2005) (emphasis added).

[56] *Id.* at 267-68 (Breyer, J., concurring).

[57] *See Batson*, 476 U.S. at 99 n.24 ("In light of the variety of jury selection practices followed in our state and federal trial courts, we make no attempt to instruct these courts how best to implement our holding today.").

[58] Wash. St. Ct. Gen. R. 37.

[59] 429 P.3d 467 (Wa. 2018).

[60] *Id.* at 470; Wash. St. Ct. Gen. R. 37(e).

nation, have resulted in the unfair exclusion of potential jurors in Washington State."[61]

This is one example in which behavioral realism about implicit bias helped a state court to reject subjective purposeful intent in favor of an objective reasonable person standard in implementing a state version of a federal constitutional guarantee.

6. State Capital Punishment

In *McCleskey v. Kemp* (1987), the United States Supreme Court was confronted with statistical evidence of gross racial disparities in capital punishment.[62] Nevertheless, it declined to find an Eighth Amendment federal constitutional violation. Revealing its continuing endorsement of the purposeful intent standard, the Court 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 ....

Nearly three decades later, the Supreme Court of the State of Washington confronted statistical evidence of racial disparities in capital punishment in *Washington v. Gregory*.[63] Since the court was applying state constitutional law, it was not strictly bound by the reasoning in *McCleskey*. Moreover, by this time, the Washington Supreme Court had embraced behavioral realism. In its opinion, for example, it noted the significance of new understandings of juvenile offenders "supported by scientific and sociological studies about the differences between juveniles and adults."[64] The court also cited cases mentioning the importance of revising law "in light of 'advances in the sci-

[61] Wash. St. Ct. Gen. R. 37(f) (emphasis added).
[62] 481 U.S. 279, 308 (1987).
[63] 427 P.3d 621 (Wash. 2018).
[64] *Id.* at 633.

entific literature'."[65] In its clearest endorsement of 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."[66]

The plaintiff's statistical analysis of racial disparities was, of course, imperfect and disputed by the state's expert. Still, the court "decline[d] to require indisputably true social science to prove that our death penalty is impermissibly imposed based on race."[67] The court instead invoked its upgraded common sense:

> Given the evidence before this court and our judicial notice of *implicit* and overt racial bias against black defendants in this state, we are confident that the association between race and the death penalty is not attributed to random chance. 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 law and history of racial discrimination provide ample support.[68]

What followed this quotation was a long string citation of authority. Among them were a prior Washington case *State v. Saintcalle* (addressing *Batson* challenges),[69] a 2011 state task force report on race and criminal justice,[70] and an amicus brief of 56 former and retired Washington state judges prepared by the ACLU of the State of Washington[71]—all of which referenced legal and empirical analyses of implicit bias. And invoking these understandings, the Washington Su-

[65] *Id.*

[66] *Id.*

[67] *Id.* at 634.

[68] *Id.* at 635.

[69] 309 P.3d 326 (Wash. 2013). *Saintcalle* cited to the task force report, which cited to legal scholarship on behavioral realism. *See, e.g.*, *Saintcalle*, 309 P.3d at 335 n.3 (citing Kang & Lane, *supra* note 18).

[70] *See* TASK FORCE ON RACE AND THE CRIMINAL JUSTICE SYSTEM, PRELIMINARY REPORT ON RACE AND WASHINGTON'S CRIMINAL JUSTICE SYSTEM 17-21 & Appendix A8 (2011).

[71] Amici Curiae Brief of 56 Former & Retired Wash. State Judges. That brief also referred frequently to "unconscious bias." *See id.* at 16-17.

preme Court struck down its death penalty under state constitutional law.

Statistics alone was not enough in 1987, but statistics *plus implicit bias* sufficed at least for the State of Washington in 2018. A similar result, for similar reasons, obtained for the State of Connecticut in the 2015 case of *Connecticut v. Santiago*.[72] These are prominent examples in which behavioral realism about implicit bias helped a state court depart from the purposeful intent requirement set in federal constitutional law when evaluating the death penalty under the state's constitution.

* * *

The above examples are just that, examples. They constitute neither the full population nor a random sample of judicial opinions, statutes, and regulations engaging with implicit social cognition or the tenets of behavior realism. There are more examples of upgraded models being mentioned in concurrences and dissents as compared to majority opinions. And there are plenty of cases in which implicit social cognition is ridiculed as junk science that should be summarily rejected. Still, in toto, the examples inventoried above tell a significant story. The scientific rise of implicit social cognition, incorporated via behavioral realism, has created meaningful elbow room for doctrinal departure away from the purposeful intent doctrine. And as legal and conceptual precedents, these cases foreshadow genuinely interesting innovations in how law can more capably respond to the complexities of discrimination.

III. Objections

The science of implicit bias threatens many on the Right because it empirically refutes the self-congratulatory claim that we are all already colorblind and that disparate impacts are caused entirely by differences in merit caused by genetics or culture. It shouldn't be that surprising, then, that President Trump's Executive Order flagged "divisive concepts" to include claims about "unconscious[]" bias. Indeed, there have been challenges to the construct (Is implicit bias really a "bias"?), measurement tools (Is an instrument like the IAT valid?), and most importantly, predictive validity (How much does implicit bias predict in

[72] See also *Connecticut v. Santiago*, 318 Conn. 1 (2015) (doing the same for Connecticut death penalty).

terms of discriminatory behavior?). To be clear, many of these challenges have been in good faith, part of the ordinary dialectic of scientific progress, which have advanced our collective understandings. Some challenges, however, feel like they've been made in less good faith, more akin to the strawperson attacks on Critical Race Theory.

Perhaps more surprising is how implicit bias has discomfited some on the Left who are uncomfortable with the epistemological privileging of science. In their view, science—especially science that is experimental, quantitative, or involves statistical null-hypothesis testing—is inevitably reductionist, acontextual, and eliding intricacies of power, history, and culture. In this vein, Jonathan Kahn worries that behavioral realists mislocate racism inside of neurons instead of larger social structures and recommend individual-focused, technocratic solutions that "not only elevate scientific authority over legal authority but also elevates unaccountable experts over officers and citizens of a democratic polity."[73]

I agree that we should resist disciplinary arrogance—the idea that knowledge or authority from any field, such as experimental social psychology, trumps all other fields in any descriptive, programmatic, or normative project. But any such reading of behavioral realism seems somewhat exaggerated and, frankly, uncharitable. Perhaps some of the Left's anxiety can be relieved by pivoting away from the question whether "science" is epistemologically exceptional to the question whether "facts" are important.

There are rich and valid critiques of "science" that come from the philosophy of science, the history of science, and science and technology studies (STS). In this chapter, I sidestep those hard debates and instead ask a simpler, more practical question: for lawyers and practitioners engaged in civil rights work, do facts matter? For instance, does it matter that a racial disparity is in fact caused by biased perception and not an accurate evaluation of the merits? If we put aside for the moment academic deconstructions of facts, objectivity, and knowledge—both playful and serious—I think the practical answer is absolutely yes. Facts matter.

Once we agree that facts matter, the next question simply becomes how do we persuade those in the undecided mid-

[73] Jonathan Kahn, Race on the Brain at 194.

dle to accept new facts, whether the subject matter be global warming, vaccines, or modern race discrimination? Even if deeply imperfect, I believe that science plays a crucial role in developing the facts. And for those on the Left who lament the use of ordinary science to promote racial equality, I ask whether they also lament the use of science to fight global warming, pandemics, and conspiracy theories.

To repeat, my basic, nearly banal point is that facts matter. Truth matters. The new truths about implicit bias are, of course, inconvenient, and as I have explicitly observed, "inconvenient truths never guarantee reform."[74] But marshaling facts is an important anti-racist strategy that is not mutually exclusive of a venerable appeal to more normative appeals to justice and fairness. To the contrary, it is an essential complement. And as I've argued, the examples above demonstrate that implicit social cognition, by way of a behavioral realism, has materially weakened our fixation on purposeful intent.

CONCLUSION

This chapter asked whether the rise of implicit social cognition had an impact on law's evolution via increased behavioral realism. The examples identified above suggest the answer is yes. I acknowledge that more cynical readings of how judges use science are, of course, possible.[75] For example, judges may be simply pursuing their political preferences, neither understanding nor genuinely caring about new facts from science. They may be invoking science only performatively—as a magician uses smoke and mirrors—to obscure what is just a naked exercise of judicial power. Although in this piece, I cannot make the argument, let me simply state my conclusion that this account sounds too jaundiced and extreme. I think that facts do matter and influence judicial thinking on the merits. Over the past two decades, judges—like the rest of us—have slowly absorbed an upgraded model of how discrimination occurs, and on this revised understanding, they have slowly bent law's arc. We may be witnessing not the beginning of the end, but the end

[74] Science article at __.

[75] Erik J. Girvan, *On Using the Psychological Science of Implicit Bias to Advance Anti-Discrimination Law*, 26 GEO. Mason U. C.R. L.J. 1, 62 (2015) (raising concerns that behavioral realism has been naive).

of the beginning, of the dominance of the purposeful intent standard for racial discrimination.