

Little Things Matter a Lot: The Significance of Implicit Bias, Practically & Legally

Jerry Kang

Skeptics point out that measures of implicit bias can only weakly predict discrimination. And it is true that under current technologies, the degree of correlation between implicit bias (for example, as measured by the Implicit Association Test) and discriminatory judgment and behavior is small to moderate. In this essay, I argue that these little effects nevertheless matter a lot, in two different senses. First, in terms of practical significance, small burdens can accumulate over time to produce a large impact in a person's life. When these impacts are integrated not only over time but double integrated over large populations, these little things become even more practically significant. Second, in terms of legal significance, an upgraded model of discrimination that incorporates implicit bias has started to reshape antidiscrimination law. This transformation reflects a commitment to "behavioral realism": a belief that the law should reflect more accurate models of human thinking and behavior.

Implicit bias is a concept that has diffused rapidly throughout our culture. One reason for the fast uptake is that it's intuitively obvious. Even without formal training in psychology or neuroscience, we realize that we navigate the world with limited cognitive resources. When confronted with a flood of sensory stimuli, what else can we do but use mental shortcuts to streamline our processing of that information. By automatically classifying any object we encounter into a category, we take advantage of our prior knowledge of and experience with that category to guide our response. For instance, if we recognize and classify something as a chair, we know how to pull it out from the table and sit down without a second thought. It doesn't matter whether that chair looks like an antique, a barstool, or an office chair, we know what a "chair" is and what to do with it. But just as we do this with chairs, we do this with people. We immediately classify a person we meet into multiple social categories, based on age, gender, race, and role. Next, meanings associated with those categories are automatically activated and guide our interaction with that person. None of this is surprising.

What is surprising is the possibility that the meanings associated with categories might be “implicit.” By implicit, I mean they are not readily subject to direct introspection. In other words, I cannot fully ascertain the meanings (that is, the attitudes and stereotypes) that I have associated with a social category by simply asking myself for an honest account. We only have partial insight into the numerous mental associations stored in our brains, which operate automatically. Even though it’s humbling to recognize that we lack perfect, introspective insight, this too isn’t exactly shocking. Every time a smell, song, or taste triggers a once-forgotten memory, we realize that traces of the past remain in our minds even if we cannot access them at will.

Finally, the recent rise of generative artificial intelligence (AI) has highlighted the computer science problem of “garbage in, garbage out.” If we train a chatbot using biased content (the garbage in), we should not be surprised that the chatbot spews biased content (the garbage out). But why would the computing machinery in our brains magically avoid this pitfall? If our own neural networks are trained through deep immersion in a social, economic, political, and media reality configured by status hierarchy, role expectations, culturally specific designations of friend versus foe, and media stereotypes, why would our brains automatically reject that learning?

In sum, one reason the concept of “implicit bias” has become so popular, so quickly, is because it makes intuitive sense. If we are honest about our limitations as thinking machines, we should not be surprised to learn that implicit biases exist and can alter our judgments. Of course, intuitive common sense is often dead wrong, so it’s important to check against the scientific evidence. Since other contributions to this issue of *Dædalus* already do so, I won’t repeat that work in detail. It suffices to say that:

1. “Implicit bias” is a valid scientific construct.
2. Implicit bias can be measured indirectly through various instruments, including reaction time measures such as the well-known Implicit Association Test (IAT).¹
3. Implicit bias is pervasive (generally favoring in-groups and those higher on a social hierarchy); related to but different from explicit bias (measured via self-reports); and generally larger in magnitude than explicit bias on socially sensitive topics such as race (and other social categories).²
4. Implicit bias predicts real-world judgment and behavior in a statistically significant way, but the effect size is small to moderate.

Numerous scientific questions remain unanswered, but outright denial of the existence of implicit bias is no longer tenable. What remains unclear is how much implicit bias matters in real-world conditions. Also uncertain are the best ways

to counter implicit bias and its consequences. The focus of this essay is to unpack what it means for implicit bias to have “small-to-moderate” effect sizes. I argue that these “little things” matter a lot, in two senses. First, in terms of practical significance, small burdens can accumulate over time to produce a large impact in a person’s life. When these impacts are integrated not only over time but double integrated over large populations, little things don’t seem so little after all. Second, in terms of legal significance, an upgraded model of discrimination based on better science, including implicit bias, has started to reshape antidiscrimination law. This happens when those who make and interpret law embrace “behavioral realism:” a belief that the law should reflect more accurate models of human thinking and behavior.

Does implicit bias have a real-world impact? More precisely, does some measure of implicit bias, produced by an instrument such as the IAT, predict real-world discrimination? For this discussion, I define “discrimination” narrowly as treating someone differently because of perceived membership in a social category, even though everyone agrees that the social category should not influence the specific decision or behavior at hand. To answer this question based on all available research (and not just cherry-picked examples), we rely on meta-analysis. A meta-analysis is an analysis of analyses. Imagine an open-source collaboration that stitches together individual snapshots taken by different photographers, using different cameras, at different times, into a single panoramic, composite picture. But instead of photos, we use academic studies. More specifically, a meta-analysis calculates a single number from all the conducted research in a domain: in this case, an “effect size” that estimates the strength of the relationship between implicit bias and intergroup discrimination.

To date, three major meta-analyses have been conducted on the predictive validity of implicit bias by researchers across the ideological spectrum.³ All meta-analyses found statistically significant effect sizes, which this literature states in terms of Pearson’s r , the correlation coefficient.⁴ The three meta-analyses, which used slightly different datasets and methodologies, calculated statistically significant correlations ranging from .10 to .24. Averaged over all three meta-analyses, the correlation is .165.⁵ By convention, this effect size is called “small-to-moderate.” To say that these correlations are “statistically significant” means roughly that they are unlikely due to chance. But most savvy readers know that *statistical* significance says little about *practical* significance.

One standard way to gauge practical significance is to square the r value to get the “percentage of variance explained.” On the simplifying assumption of uniform variability in the r values measured across the meta-analyses, we get the r^2 value of .027 (.165 \times .165 = .027). In other words, implicit bias would explain 2.7 percent of the total variance (a statistical term of art) measured in the intergroup

behavior. Your immediate reaction, even if you can't recall the statistical definition of variance, might be that this seems like a small percentage. Perhaps it's too small an effect for us to care about.

Indeed, this precise objection has long been raised by skeptical academics and advocates. For example, in 2005, legal scholar Amy Wax and psychologist Philip Tetlock editorialized in *The Wall Street Journal* that "there is often no straightforward way to detect discrimination of any kind, let alone discrimination that is hidden from those doing the deciding."⁶ In 2009, Tetlock and legal scholar Gregory Mitchell worried that implicit bias researchers were politicizing science and raised objections to believing that implicit bias caused discrimination in the real world.⁷ In their 2015 meta-analysis, psychologist Frederick Oswald and colleagues (including Tetlock and Mitchell) lamented that "researchers still cannot reliably identify individuals or subgroups . . . who will or will not act positively, neutrally, or negatively toward members of any specific in-group or out-group."⁸

In the legal domain, consider also the dismissive attitude reflected in court opinions rejecting the expert testimony of psychologist Anthony G. Greenwald, who invented the Implicit Association Test:

- "The application of Dr. Greenwald's cognitive theory on stereotyping to the circumstances at the Y[MCA] is speculative, *without any scientific basis*."⁹
- "This sort of superficial analysis . . . is not expert material; it is *the say-so of an academic who assumes* that his general conclusions from the IAT would also apply to [the defendant]."¹⁰

These examples demonstrate that the question of practical significance indeed remains a live controversy. How then should we think about the problem of small effect sizes?

First, we should not assume that small, measured r values are necessarily worthless. Back in 1985, cognitive psychologist Robert Abelson made this point powerfully with a baseball analogy. He asked, "What percentage of the variance in athletic outcomes can be attributed to the skill of the players, as indexed by past performance records?"¹¹ In simpler terms, how much does a typical player's batting skill (measured by batting average) explain the percentage of variance in any single at bat? The answer turned out to be spectacularly low: approximately one-third of 1 percent, which is equivalent to an effect size of $r = .056$.¹² Recall that the effect size measured for implicit bias was almost three times larger at $r = .165$. Even if we compared players whose batting averages were two standard deviations above the mean, to those who were two standard deviations below (roughly a .320 hitter compared to a .220 hitter), for a single at bat, skill would explain only 1.3 percent of the variance, which is equivalent to $r = .113$. With two outs in the final inning and a player in scoring position, every manager would re-

place a .220 hitter scheduled to bat with a .320 pinch hitter if available. But this reveals that a small-to-moderate effect size of $r = .113$ is practically significant in the multimillion-dollar sport of professional baseball.

Second, even if any single instance of discrimination caused by implicit bias seems trivial (such as a misperception or less friendly body language), we must consider their accumulation over time. Abelson explained his surprising findings by pointing out that batting skill manifests over multiple at bats during an entire season. As psychologists David Funder and Daniel Ozer elaborate:

The typical Major League baseball player has about 550 at bats in a season, and the consequences cumulate. This cumulation is enough, it seems, to drive the outcome that a team staffed with players who have .300 batting averages is likely on the way to the playoffs, and one staffed with players who have .200 batting averages is at risk of coming in last place. The salary difference between a .200 batter and a .300 batter is in the millions of dollars for good reason.¹³

All this should remind us of the phrase “death by a thousand paper cuts.” To integrate all implicit bias-actuated harms over time, we need to know frequency (how many “cuts” per unit of time) and duration (what time period to measure from beginning to end). Depending on the question, duration can be years at a firm, in an industry, in a career, and indeed one’s entire lifetime. And frequency is not just one critical judgment every few years when we apply for a job or promotion. Instead, it could be every social, economic, political, and professional interaction. It could be every time we get into a parking dispute; every time we get pulled over for a traffic stop; every time we ask for help at a hardware store; every time we shop for furniture, a car, or a house; every time we apply for a credit card or loan; every time we wait to be seated at a restaurant; every time we apply for a job or promotion; every time we turn in mediocre work and get (or don’t get) the benefit of the doubt; every time we join a team; every time credit is shared; every holiday office party; and so on. In some sense, the frequency is multiple times per day because almost no social interaction is immune from implicit biases. This amounts to far more than a thousand cuts.

Third, after integrating “paper cuts” across time to assess an individual’s harm, we should double integrate over all people potentially affected. An illuminating example comes from public health. Back in 1990, psychologist Robert Rosenthal pointed out that in a clinical trial, scientists noticed a statistically significant correlation of $r = .034$ between taking aspirin and reduced chances of heart attack.¹⁴ Even though the correlation was small (almost five times smaller than the effect size for implicit bias), the scientists stopped the randomized double-blind study because they felt it was not ethical to continue giving the control group placebos. In Greenwald’s account of that study, he considered the population-level impacts of decreasing the chances of heart attack even marginally for each participant.¹⁵

Given the millions of people subject to heart attack, aspirin could prevent approximately four hundred and twenty thousand heart attacks over a five-year period – something we all presumably agree is practically significant. Considering these three lessons – the batting average, a thousand paper cuts, and double integrals across time and people – will produce a more thoughtful understanding of practical significance. Still, having more concrete examples is helpful, and one way to produce them is to run simulations under plausible assumptions.

For example, Greenwald and psychologists Mahzarin R. Banaji and Brian Nosek modeled the potential impact of implicit bias on racial profiling. Suppose that implicit bias nudges police officers to cite Black drivers and pedestrians more frequently than White ones. Assuming that the effect size was just $r = .148$ (a value calculated in one of the Oswald meta-analyses highly critical of implicit bias), Greenwald and colleagues imagined two different worlds. In World 1, all the police officers were one standard deviation lower on implicit bias, and in World 2, all the police officers were one standard deviation higher on implicit bias. If we compared these two worlds, World 1 would have 9,976 fewer Black stops, which amounted to 5.7 percent of the total number of stops for the year of data analyzed.¹⁶ Who would argue that avoiding nearly ten thousand police stops of Black people annually is practically insignificant?

In another example, Greenwald created a simulation to estimate how much implicit bias could alter the expected prison sentence for committing a crime. With plausible assumptions (a crime with a mean sentence of five years and a standard deviation of two years), implicit bias effect size of $r = .10$, and a five-round model (involving arrest, arraignment, plea bargain, trial, and sentencing), the simulation found that a Black criminal can expect a probabilistic sentence of 2.44 years versus a White criminal expecting 1.40 years. Remember that we must integrate this individual-level differential over the entire relevant population of criminal cases in any given year, which can run into the tens of thousands.¹⁷ Even if there were only one thousand cases of this sort per year, implicit bias would produce one thousand years of more Black imprisonment annually. Again, how can this be practically insignificant?

Consider one last simulation involving Big Law. Assume that, to make partner, litigation associates must survive a monthly up-or-out tournament that lasts for eight years. Suppose that implicit bias creates just a 1 percent difference in the monthly survival rate, with the White associate likely to survive at 99 percent but the Asian associate likely to survive at 98 percent.¹⁸ For simplicity's sake, if we assume each month's survival rate to be an independent probability, the White associate's chances of making partner (which requires surviving $8 \times 12 = 96$ cuts) would be 38.1 percent, whereas the chances for the Asian associate would be 14.4 percent.¹⁹ And what is the r value equivalent for that 1 percent difference in

monthly survival rate? It amounts to a mere $r = .04$. The reason why such a small correlation can produce such drastic results is because each month a critical decision (up or out) is being made, and we are considering the accumulated impact of such decisions over ninety-six months.

There are bones to pick, of course, with the above simulations as being too stylized and not realistic. One can also object that the predictive validity studies were not conducted out in the field, under real-world circumstances, which include legal and procedural checks on discrimination. These are fair criticisms. But when we insist on greater realism, better evidence, or larger effect sizes, we should do so consistently, without double standards. For example, let's compare implicit bias to medical phenomena that we generally accept as practically significant. We already made one such comparison with aspirin and heart attacks. In 2001, psychologist Gregory Meyer and colleagues compiled a useful inventory of the effect sizes of what might be called medical "common sense."²⁰ Interestingly, they were often lower than or on par with the effect size found for implicit bias ($r = .165$):

- antihypertensive medication and reduced risk of stroke ($r = .03$),
- chemotherapy and surviving breast cancer ($r = .03$),
- antibiotic treatment of acute middle ear pain in children and improvement within seven days ($r = .08$),
- alcohol use during pregnancy and subsequent premature birth ($r = .09$),
- combat exposure in Vietnam and subsequent PTSD within eighteen years ($r = .11$),
- extent of low-level lead exposure and reduced childhood IQ ($r = .12$),
- nonsteroidal anti-inflammatory drugs and pain reduction ($r = .14$),
- post-high school grades and job performance ($r = .16$),
- validity of employment interviews for predicting job success ($r = .20$), and
- effect of alcohol on aggressive behavior ($r = .23$).

Given these comparable effect sizes, will those who object to the practical significance of implicit bias similarly object to the practical significance of these other phenomena? Second, let's compare the practical significance of implicit bias with that of explicit bias. When we discover that someone has explicit bias, we typically take note. For example, when meeting a new neighbor, if they blurt out anti-Semitic tropes, we will presumably take note. Similarly, during *voir dire* (the process of questioning potential jurors), if someone expresses stereotypes that Latinos are culturally prone to criminal gang activity, we will again take note. When we notice such expressions of explicit bias, we don't chastise ourselves for being irrational, credulous, "woke," or ideological. But here's where things get in-

teresting: the same meta-analyses that found small-to-moderate effect sizes for implicit bias revealed that implicit bias scores have comparable or *more* predictive power than explicit bias scores.²¹ This suggests that if we take explicit bias seriously (because it might predict discriminatory judgment and behavior), we should take implicit bias even more seriously.

Third, let's compare the effect size of implicit bias with effect sizes that are often deemed legally significant in civil rights enforcement. The Equal Employment Opportunity Commission's (EEOC) Uniform Guidelines on Employee Selection Procedures adopt a rule of thumb that when a selection rate for any protected category is less than four-fifths of the rate for the group with the highest success rate, this disparity will be regarded as *prima facie* evidence of adverse impact, which is the first step of winning a disparate impact case under Title VII of the 1964 Civil Rights Act.²² What does this rule of thumb mean in terms of effect sizes? Consider the following hypothetical about junior-level promotions in a national firm.

Among White applicants in any given year, suppose that five hundred are promoted and five hundred are not. In other words, the promotion rate for White people is 50 percent (five hundred out of one thousand total). Next, suppose that among Asian applicants (a smaller population), thirty-nine are promoted and sixty-one are not; the promotion rate is thus 39 percent (thirty-nine out of one hundred total). Because the ratio of promotion rates (39 percent Asian to 50 percent White) is lower than four-fifths, agency guidelines instruct judges to find *prima facie* evidence of a disparate impact. What do these differences in promotion rates look like when they are converted into Pearson's r ? The $r = .063$. In other words, the federal government has announced a rule of thumb suggesting legal significance – under plausible assumptions of population size and promotion rates – for an effect size that is only $r = .063$. On what grounds, then, can we reflexively dismiss implicit bias ($r = .165$) as practically insignificant?

In sum, a careful inquiry into practical significance reveals that phenomena with small effect sizes can be practically significant. Little things mean a lot, not only in the trajectory of individual lives but also in the arc of entire peoples. In addition, we should actively scan for double standards. For example, if we happily rely on medical common sense – as we pop supplements, avoid heart attacks, or decide on treatment for breast cancer – we should recognize that we do so often because of r values *lower* than the effect sizes found with implicit bias. If we dismiss implicit bias as practically insignificant, then what justifies the double standard in our own self-care? Could it be that we worry about our own health and beauty but not so much about implicit bias – mediated harms inflicted on others?

Also, if we care so deeply about explicit bias, enough to interrogate potential jurors about their prejudices and stereotypes publicly during *voir dire*, on what sci-

entific grounds should we dismiss implicit bias as unimportant? To recap, over the past three decades in the mind sciences, researchers have uncovered surprising evidence that discrimination may be caused by implicit bias. How should these new discoveries influence the law? For two decades, I have advocated for a school of thought called “behavioral realism,” which combines the traditions of legal realism and behavioral science. Stated succinctly, behavioral realism insists that law should incorporate more realistic models of human behavior.

This approach involves a three-step process.²³ First, we should regularly scan the sciences for more accurate, upgraded models of human decision-making and behavior. Second, we should compare that upgraded model to the “commonsense” legacy understandings embedded within the current law. Third, when the gap between the upgraded and legacy models grows sufficiently large (however defined), we should revise the law or its interpretation in accordance with the upgraded model. If that can’t be done – for example, because of controlling precedent, constitutional constraints, or other overriding moral or policy considerations – then lawmakers should clearly explain their reasons why.²⁴ This requirement applies to judges and administrative agencies, in particular, who are obliged to give reasons for how they interpret and make law.²⁵ This simple three-step process largely avoids contentious normative questions and instead draws on a broadly overlapping consensus regarding 1) promoting instrumental rationality and 2) avoiding hypocrisy.

Concerning instrumental rationality, importing an upgraded, more behaviorally realistic model of decision-making means that the law will function under more accurate descriptions of human action. Doing so will be more efficient. For example, if the mind sciences discover better ways to deter bad behavior in adolescents, white-collar criminals, and large corporations, it would be instrumentally rational to incorporate these insights into our legal deterrence regimes. Concerning hypocrisy, all laws, including antidiscrimination laws, have some publicly announced purpose. When we learn that their purpose cannot be well-achieved because we are relying on legacy understandings, we should do something about it. If we decline to do so without good reason, we risk hypocrisy. For example, suppose a bank adopted cybersecurity measures – such as firewalls, multifactor authentication, and password managers – to prevent online fraud and other security breaches. But the bank discovers that its measures have failed all along because they fundamentally misunderstood underlying vulnerabilities like social engineering. If the bank declines to adapt to this realization, can we believe that it cares about security? And if it continues to tout its commitment to security, would we not criticize such advertising as deluded or hypocritical? As I elaborate below, this simple approach of behavioral realism has already started to influence antidiscrimination law.

A central feature of American civil rights law is the stylized distinction between intentional discrimination and disparate impact. On one hand, most antidiscrimination laws require a showing of intentional discrimination, which generally means that the defendant purposefully treated someone differently because of their social category. The focus is on the mental state of the individual defendant and their deliberate, purposeful consideration of a social category. On the other hand, some civil rights laws require only a showing of disparate impact.²⁶ As long as a specific practice causes a disparate impact across legally protected social categories, that practice must be specially justified.

In the employment context, a disparate impact-causing practice must be functionally necessary in the sense that it must be job-related and a business necessity. In addition, if there is an alternative policy or practice that produces equally good results with less disparate impact, the defendant will be held liable if they refuse to adopt it. The focus of disparate impact liability is not on the individual defendant's state of mind; instead, it is on group consequences. Even without legal training, one can see how disparate impact theory casts a broader net for legal concern than intentional discrimination. After all, many facially neutral selection criteria, adopted and applied without purposeful intentional discrimination, can produce a disparate impact.

For example, if there is an average height difference between Asian Americans and White Americans, then a minimum height requirement for first responders – originally adopted and applied without consideration of race – can produce a disparate racial impact. It was precisely this anxiety of disparate impact overreach that led the Supreme Court to read the federal Constitution's Equal Protection Clause narrowly, to proscribe only intentional discrimination. The historic case was *Washington v. Davis* (1976).²⁷ 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. The court's policy rationale was explicit:

A rule that a statute designed to serve neutral ends is nevertheless invalid, absent compelling justification, if in practice it benefits or burdens one race more than another would be far-reaching and would raise serious questions about, and perhaps invalidate, a whole range of tax, welfare, public service, regulatory, and licensing statutes that may be more burdensome to the poor and to the average black than to the more affluent white.²⁸

Intentional discrimination remains the constitutional touchstone and the initial presumption in interpreting all antidiscrimination laws. Moreover, as noted above, "intentional" is often presumed to mean "purposeful" and not a lower lev-

el of culpability, such as “knowing,” “reckless,” or “negligent.”²⁹ Unfortunately, proving that the defendant purposefully treated someone worse because of a protected social category is extraordinarily difficult. That’s why Critical Race Theorists have criticized the intentional discrimination requirement as privileging the “perpetrator perspective.”³⁰ Has the science of implicit bias, by way of behavioral realism, weakened this fixation? Consider the following examples.

In *Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc.* (2015),³¹ the Supreme Court had to interpret the federal Fair Housing Act (FHA), which declares it unlawful “to refuse to sell or rent, or otherwise make unavailable . . . a dwelling to any person because of race [and other protected categories].”³² The question presented was whether the statute required purposeful intentional discrimination, or might it also recognize disparate impact? In a 5–4 decision, the court recognized a disparate impact theory of liability. Per Justice Anthony Kennedy:

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.³³

According to the court, the more capacious disparate impact theory of liability was better suited to respond to “unconscious prejudices.” Partly because the court accepted an upgraded model of human decision-making, which included the possibility of discrimination based on implicit social cognitions, the court adopted a broader interpretation of the Fair Housing Act to include disparate impact liability.

In *Kimble v. Wisconsin Department of Workforce Development* (2010), the Eastern District of Wisconsin heard an employment discrimination case under Title VII of the 1964 Civil Rights Act.³⁴ Title VII recognizes both disparate treatment (intentional discrimination) and disparate impact theories of liability. For disparate treatment, courts frequently suggest that the defendant must have explicitly and purposefully used a protected social category in its decision-making. But in truth, the statute does not specify any such mental state. Instead, it simply prohibits employment discrimination “because of” a person’s race and other protected social categories. With this textual flexibility in mind, the court pivoted away from purposeful intent and instead asked more literally for category causation.³⁵ 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.”³⁶ Applying this clarified legal understanding to the facts of the

case, 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*.”³⁷ It noted that because of the ordinary psychological process of categorical thinking, a supervisor may use stereotypes “*whether or not* the supervisor is *fully aware* that this is so.”³⁸ Again, an upgraded model of discrimination, which the court gleaned in part from secondary sources advocating behavioral realism, led the court to rule in favor of the plaintiff.³⁹

In *State v. Gill* (2019),⁴⁰ the Court of Appeals for Kansas had to interpret a state statute that prohibited “racial or other biased-based policing.”⁴¹ The case was prompted by a police officer approaching two Black men in an SUV because they were allegedly “staring hard” at him, which resulted in a search that uncovered drugs. The trial court found a statutory violation, and on appeal, the appellate court affirmed. The dissent railed loudly at the majority for “brand[ing] an officer of the law . . . a racist . . . [without] evidence supporting such a serious charge.”⁴² But importing an upgraded, more behaviorally realistic model of discrimination, the majority de-escalated and explained that “no one here is branding [the officer] a racist.”⁴³ Instead, the relevant question was one of racial causation, whether the officer “let racial bias – *conscious or unconscious* – affect his initiation of enforcement action.”⁴⁴

In *Woods v. City of Greensboro* (2017),⁴⁵ the Fourth Circuit Court of Appeals reviewed a district court’s granting of a motion to dismiss a 42 U.S.C. § 1981 civil rights action (equal contracting rights) for failure to state a claim. The appellate court 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.”⁴⁶ 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*.”⁴⁷

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.”⁴⁸ For these reasons, the Court of Appeals reversed the dismissal and allowed the case to proceed to discovery.

The Supreme Court of the State of Washington deserves special recognition as a trailblazer for behavioral realism. Consider, for example, how it has evolved the processing of peremptory challenges. Way back in *Batson v. Kentucky* (1986), the United States Supreme Court held that a prosecutor’s pur-

purposeful discrimination to strike jurors because of race violated federal equal protection guarantees.⁴⁹ Unfortunately, it was nearly impossible to prove such a state of mind because any competent prosecutor could provide non-race-based justifications for striking a potential juror.

In 2018, the Washington Supreme Court pivoted away from demanding proof of a prosecutor's *subjective* mental state. Instead, the court adopted an *objective* reasonable person standard via judicial rulemaking (General Rule 37) and opinion in *Washington v. Jefferson* (2018).⁵⁰ Their revised approach asks whether "an *objective observer* could view race or ethnicity *as a factor* in the use of the peremptory challenge."⁵¹ What's fascinating is that this objective observer benefits from a fully upgraded model of discrimination. General Rule 37(f) expressly states: "For purposes of [the Nature of Observer] rule, an objective observer is aware that *implicit, institutional, and unconscious biases*, in addition to purposeful discrimination, have resulted in the unfair exclusion of potential jurors in Washington State."⁵²

Other states have followed Washington's lead. For example, in 2020, the California legislature passed AB 3070, which targeted "the use of group stereotypes and discrimination, whether based on *conscious or unconscious bias*, in the exercise of peremptory challenges."⁵³ California's statute does not require proof of intentional discrimination; instead, upon a challenge, the court must determine 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 . . ."⁵⁴

In 2021, the Arizona Supreme Court eliminated all peremptory challenges in part due to the problem of implicit bias.⁵⁵ In 2022, upon the recommendations of a judicial task force, the judges of Connecticut's Superior Court amended their Practice Book not to require any showing of purposeful discrimination. Instead, 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."⁵⁶ 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."⁵⁷

Finally, in 2022, New Jersey's Supreme Court amended its Rules Governing the Courts of 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.⁵⁸ The Official Comment lists reasons that are presumptively invalid because they are historically associated with "improper discrimination, explicit bias, and *implicit bias*."⁵⁹

Consider also how the Washington Supreme Court diverged from the path created by *McCleskey v. Kemp* (1987).⁶⁰ In *McCleskey*, the United States Supreme Court

declined to find an Eighth Amendment federal constitutional violation based on statistical evidence showing gross racial disparities in capital punishment. 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.⁶¹

Nearly three decades later, in *Washington v. Gregory* (2018), the Washington Supreme Court explained the importance of revising law “in light of ‘advances in the scientific literature.’”⁶² 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.”⁶³ The court then alloyed statistical evidence of racial disparities in capital punishment with an upgraded psychological model of discrimination to find a state constitutional violation:

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.⁶⁴

Although statistics alone were not enough in 1987 for the federal Supreme Court, statistics coupled with general awareness of implicit bias sufficed for the state of Washington in 2018.⁶⁵ As the above examples demonstrate, by embracing behavioral realism, courts have imported more accurate models of discrimination that account for implicit bias. And through these upgraded understandings, courts have interpreted and applied both substantive and procedural laws differently than they would have under legacy beliefs. These cases evince the legal significance of implicit bias.

To be clear, these examples speak more to future potential than current actualization. As pointed out above, in the discussion of effect sizes, there are many courts that dismiss implicit bias as politicized, exaggerated, inflammatory, and too general to help decide specific cases. In addition, the censorship of so-called dangerous ideas, such as Critical Race Theory and implicit bias, will exact its political toll. But as also demonstrated above, we have already witnessed significant examples of legal transformation based on the evidence of implicit bias.

Intriguingly, the Supreme Court’s recent, aggressive turn toward “but-for” causation in antidiscrimination law may spawn still more opportunity. In *Comcast Corporation v. National Association of African American-Owned Media, et al.* (2020), the Supreme Court adopted a baseline understanding based on “‘textbook tort law’ that . . . a plaintiff must demonstrate that, but for the defendant’s unlawful conduct, its alleged injury would not have occurred. . . . That includes when it comes to federal anti-discrimination laws”⁶⁶ The court’s objective was to restrict “mixed motive” cases, in which the plaintiff could prevail on a discrimination claim if race (or some other protected social category) was one “motivating factor” among many, even if it were not the “but-for” cause.

Consider the unexpected opportunity that this standard creates, however, for incorporating implicit bias.⁶⁷ If we take “but-for” causation seriously, that means we ask a simple counterfactual question: if the Black person were White, would they have been treated the same? We do not have to make findings about purposeful intent or whether the defendant subjectively and self-consciously considered race, which is so hard to prove. Instead, we are simply left with a probabilistic question of fact, about “but-for” causation, to be decided by the fact finder, based on all admissible evidence and their model of human decision-making.

The science of implicit bias is paradoxically both intuitive and disorienting. On the one hand, we know that our brain leverages schemas and categories to efficiently process the world, and the fact that we might do so with human social categories should not surprise us. On the other hand, because we have been taught that discrimination is wrong, it disorients us to find out that we may be discriminating without even realizing. A natural defensive reaction is to simply dispute the science as incorrect. When outright denial is impossible, given that the findings are statistically significant, the next step is to minimize the harm and deny their practical significance because of low effect sizes. As I have demonstrated, however, little things matter a lot. And if we resist double standards, we see that implicit bias is indeed a matter of practical significance for individuals and for society.

These new facts about implicit social cognition have provided us with a more behaviorally realistic model of discrimination. This upgraded model has rapidly diffused throughout our culture and has made inroads even into the staid law. It would be naive to assume that by virtue of greater accuracy and realism the model will necessarily prevail. Surely politics and ideologies will have their say. But over the past quarter-century, the evolving science of implicit bias has presented us with a stark choice. We can act like ostriches, burying our heads in the sand, and selectively insist on metaphysical certitude before taking corrective action. Or we can concede our cognitive limitations, roll up our sleeves, and try to design better policies, procedures, practices, and even laws to prevent discrimination from its various causes – including implicit bias.

ABOUT THE AUTHOR

Jerry Kang is Distinguished Professor of Law and Distinguished Professor of Asian American Studies (by courtesy) at the University of California, Los Angeles. He is the author of *Communications Law and Policy: Cases and Materials* (with Alan Butler and Blake E. Reid, 2023) and *Race, Rights, and Reparation: Law and the Japanese American Internment* (with Eric K. Yamamoto, Margaret Chon, Carol L. Izumi, and Frank H. Wu, 2013).

ENDNOTES

- ¹ See Anthony G. Greenwald and Calvin K. Lai, “Implicit Social Cognition,” *Annual Review of Psychology* 71 (1) (2020): 419, 423–424, providing three categories of instruments: 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. For descriptions of the IAT, see Kate A. Ratliff and Colin Tucker Smith, “The Implicit Association Test,” *Dædalus* 153 (1) (Winter 2024): 51–64, <https://www.amacad.org/publication/implicit-association-test>. In the legal literature, see Jerry Kang and Kristin Lane, “Seeing through Colorblindness: Implicit Bias and the Law,” *UCLA Law Review* 58 (2) (2010): 465, 472–473; and Kristin A. Lane, Jerry Kang, and Mahzarin R. Banaji, “Implicit Social Cognition and the Law,” *Annual Review of Law and Social Science* 3 (1) (2007): 427, 428–431.
- ² Kirsten N. Morehouse and Mahzarin R. Banaji, “The Science of Implicit Race Bias: Evidence from the Implicit Association Test,” *Dædalus* 153 (1) (Winter 2024): 21–50, <https://www.amacad.org/publication/science-implicit-race-bias-evidence-implicit-association-test>; Kate A. Ratliff, Nicole Lofaro, Jennifer L. Howell, et al., “Documenting Bias from 2007–2015: Pervasiveness and Correlates of Implicit Attitudes and Stereotypes II” (unpublished pre-print); and Brian A. Nosek, Frederick L. Smyth, Jeffrey J. Hansen, et al., “Pervasiveness and Correlates of Implicit Attitudes and Stereotypes,” *European Review of Social Psychology* 18 (1) (2007): 36–88.
- ³ Anthony G. Greenwald, T. Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R. Banaji, “Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity,” *Journal of Personality and Social Psychology* 97 (1) (2009): 17, 19–20, explaining that $r = .24$ for Black/White bias; Frederick Oswald, Gregory Mitchell, Hart Blanton, et al., “Predicting Ethnic and Racial Discrimination: A Meta-Analysis of IAT Criterion Studies,” *Journal of Personality and Social Psychology* 105 (2) (2013): 171–192, explaining that $r = .15$ on Black/White implicit bias; and Benedek Kurdi, Allison E. Seitchik, Jordan R. Axt, et al., “Relationship Between the Implicit Association Test and Intergroup Behavior: A Meta-Analysis,” *The American Psychologist* 74 (5) (2019): 569–586.
- ⁴ The correlation coefficient indicates the strength of the linear relationship between two variables, in this case implicit bias and intergroup behavior. If the relationship is perfectly linear, then $r = \pm 1.0$, where a +1 value indicates a perfectly positive linear relationship and a –1 indicates a perfectly negative linear relationship. A value of $r = .0$ would indicate that there is no linear relationship between the two variables.
- ⁵ Anthony G. Greenwald, Nilanjana Dasgupta, John F. Dovidio, et al., “Implicit-Bias Remedies: Treating Discriminatory Bias as a Public-Health Problem,” *Psychological Science in the Public Interest* 23 (1) (2022): 7, 11.
- ⁶ Amy Wax and Philip E. Tetlock, “We Are All Racists at Heart,” *The Wall Street Journal*, December 1, 2005, <https://www.wsj.com/articles/SB113340432267610972>.

- ⁷ See Gregory Mitchell and Philip E. Tetlock, “Antidiscrimination Law and the Perils of Mind Reading,” *Ohio State Law Journal* 67 (1) (2006): 1023, 1056, identifying concerns about “internal validity” (causation) and “external validity” (applicability, real-world circumstances). For a reply to this “junk science” critique, see Kang and Lane, “Seeing through Colorblindness,” 504–509.
- ⁸ Frederick L. Oswald, Gregory Mitchell, Hart Blanton, et al., “Using the IAT to Predict Ethnic and Racial Discrimination: Small Effect Sizes of Unknown Societal Significance,” *Journal of Personality and Social Psychology* 108 (4) (2015): 562, 569.
- ⁹ *Jones v. Nat’l Council of YMCA*, 2013 WL 7046374, *9 (N.D. Ill. 2013) (report of Magistrate Judge Arlander Keys).
- ¹⁰ *Karlo v. Pittsburgh Glass Works, LLC*, 2015 WL 4232600, *7 (W.D. Penn. 2015), *affirmed*, 849 F.3d 61 (3d Cir. 2017) (affirming on narrower grounds).
- ¹¹ Robert P. Abelson, “A Variance Explanation Paradox: When a Little is a Lot,” *Psychological Bulletin* 97 (1) (1985): 129, 129–130.
- ¹² See *ibid.*, 131, reporting variance explained as .00317, the square root of which is equivalent to $r = .056$.
- ¹³ David C. Funder and Daniel J. Ozer, “Evaluating Effect Size in Psychological Research: Sense and Nonsense,” *Advances in Methods and Practices in Psychological Science* 2 (2) (2019): 156, 161.
- ¹⁴ See Robert Rosenthal, “How Are We Doing in Soft Psychology?” *The American Psychologist* 45 (6) (June 1990): 775.
- ¹⁵ See Anthony G. Greenwald, Mahzarin R. Banaji, and Brian A. Nosek, “Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects,” *Journal of Personality and Social Psychology* 108 (4) (2015): 553, 558.
- ¹⁶ *Ibid.*, 558.
- ¹⁷ See Jerry Kang, Mark Bennett, Devon Carbado, et al., “Implicit Bias in the Courtroom,” *UCLA Law Review* 59 (5) (2012): 1124, 1151.
- ¹⁸ For evidence of an implicit stereotype in favor of White men versus East Asian men as litigators and how it influences the evaluation of cross-examinations, see Jerry Kang, Nilanjana Dasgupta, Kumar Yogeeswaran, and Gary Blasi, “Are Ideal Litigators White? Measuring the Myth of Colorblindness,” *Journal of Empirical Legal Studies* 7 (4) (2010): 886, 900–906.
- ¹⁹ See Jerry Kang, “What Judges Can Do About Implicit Bias,” *Court Review* 57 (2) (2021): 78, 80–81.
- ²⁰ See Gregory J. Meyer, Stephen E. Finn, Lorraine D. Eyde, et al., “Psychological Testing and Psychological Assessment: A Review of Evidence and Issues,” *The American Psychologist* 56 (2) (2001): 128, 130 (Table 1).
- ²¹ See Greenwald, Poehlman, Uhlmann, and Banaji, “Understanding and Using the Implicit Association Test,” 73 (Table 3), finding that implicit attitude scores predicted behavior in the Black/White domain at an average correlation of $r = .24$, whereas explicit attitude scores had correlations of average $r = .12$. See also Kurdi, Seitchik, Axt, et al., “Relationship Between the Implicit Association Test and Intergroup Behavior,” 569–586, finding that implicit biases provide a unique contribution to predicting behavior ($\beta = .14$) and does so more than explicit measures ($\beta = .11$).

- ²² The Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. §1607.4(D) (2023).
- ²³ See, for example, Jerry Kang, “Rethinking Intent and Impact: Some Behavioral Realism about Equal Protection,” *Alabama Law Review* 66 (3) (2015): 627–651 (2014 Meador Lecture on Equality); and Jerry Kang, “The Missing Quadrants of Antidiscrimination: Going Beyond the ‘Prejudice Polygraph,’” *Journal of Social Issues* 68 (2) (2012): 314–327.
- ²⁴ For a fuller account, see Kang and Lane, “Seeing through Colorblindness,” 490–492.
- ²⁵ This essay does not discuss how judges should try to avoid implicit bias in their own decision-making. For analysis and recommendations, see Kang, “What Judges Can Do about Implicit Bias,” 78–91.
- ²⁶ See *Griggs v. Duke Power Co.*, 401 U.S. 424, 431 (1971) (introducing disparate impact theory for Title VII employment discrimination). This theory of liability was later ratified by Congress in 1991. Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071, 1074 (codified at 42 U.S.C. § 2000e-2[k]).
- ²⁷ See *Washington v. Davis*, 426 U.S. 229 (1976).
- ²⁸ *Ibid.*, 248. This interpretation was strengthened three years later in *Personnel Adm’r of Massachusetts v. Feeney*, 442 U.S. 256 (1979), in the context of gender.
- ²⁹ I take these gradations from the Model Penal Code § 2.02 (General Requirements of Culpability). *Purposely* means 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.” *Ibid.* [emphasis added]. In the Model Penal Code, “intentionally” or “with intent” means purposely. See MPC § 1.13(12) (Definitions) (“‘intentionally’ or ‘with intent’ means purposely”).
- ³⁰ See Alan David Freeman, “Legitimizing Racial Discrimination Through Antidiscrimination Law: A Critical Review of Supreme Court Doctrine,” *Minnesota Law Review* 62 (6) (1978): 1049–1119.
- ³¹ *Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc.*, 576 U.S. 519 (2015).
- ³² 42 U.S.C. § 3604(a), § 3605(a).
- ³³ *Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc.*, (2015), 540 [emphasis added].
- ³⁴ *Kimble v. Wisconsin Department of Workforce Development*, 690 F. Supp. 2d 765 (E.D. Wis. 2010).
- ³⁵ For early scholarship recommending a category causation standard, see Linda Hamilton Krieger and Susan T. Fiske, “Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment,” *California Law Review* 94 (4) (2006): 997, 1053–1054; and Linda Hamilton Krieger, “The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity,” *Stanford Law Review* 47 (6) (1995): 1161, 1226.

- ³⁶ See 690 F.Supp.2d 765, 768–769 [emphasis added].
- ³⁷ *Ibid.*, 775–776 [emphasis added].
- ³⁸ *Ibid.*, 776 [emphasis added].
- ³⁹ See *ibid.*, 776 (citing articles appearing in the 2006 Behavioral Realism Symposium).
- ⁴⁰ *State of Kansas v. Davon M. Gill*, 56 Kan. App. 2d 1278 (2019).
- ⁴¹ Kan. Stat. Ann. § 22-4609.
- ⁴² 56 Kan. App. 2d 1278, 1288 (Powell, J., Dissenting).
- ⁴³ *Ibid.*, 1286.
- ⁴⁴ *Ibid.*, 1286–1287 [emphasis added].
- ⁴⁵ *Woods v. City of Greensboro*, 855 F.3d 639 (4th Cir. 2017).
- ⁴⁶ *Ibid.*, 641 [emphasis added].
- ⁴⁷ *Ibid.*, 651–652 [emphasis added].
- ⁴⁸ *Ibid.*, 652 [emphasis added].
- ⁴⁹ *Batson v. Kentucky*, 476 U.S. 79 (1986).
- ⁵⁰ Wash. St. Ct. Gen. R. 37; and *Washington v. Jefferson*, 429 P.3d 467 (Wa. 2018).
- ⁵¹ *Ibid.*, 470. See also Wash. St. Ct. Gen. R. 37(e).
- ⁵² Wash. St. Ct. Gen. R. 37(f) [emphasis added].
- ⁵³ Assem. Bill 3070, ch. 318 (Cal. 2020), codified at Cal. Civ. Proc. Code § 231.7, Sec. 1(c) [emphasis added].
- ⁵⁴ Cal. Civ. Proc. Code § 231.7, Sec. 2(d) [emphasis added].
- ⁵⁵ See Order Amending Rules 18.4 and 18.5 of the Rules of Criminal Procedure, and Rule 47(e) of the Rules of Civil Procedure, No. R-21-0020 (Ariz. 2021). Subsequently, there was a legislative effort in Arizona to reinstate peremptory challenges in criminal cases, but this attempt was rejected by Arizona’s legislature. See H.B. 2413 (Ariz. 2022).
- ⁵⁶ “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 [emphasis added].
- ⁵⁷ “Sec. 5–12(e),” in *Official 2023 Connecticut Practice Book*, 180 [emphasis added].
- ⁵⁸ N.J. Ct. R. 1:8–3A [emphasis added].
- ⁵⁹ *Ibid.*, Official Comment (3) [emphasis added].
- ⁶⁰ *McCleskey v. Kemp*, 481 U.S. 279 (1987).
- ⁶¹ *Ibid.*, 312–313.
- ⁶² *Washington v. Gregory*, 427 P.3d 621, 633 (Wash. 2018) (quoting *State v. O’Dell*, 358 P.3d 359 [2015]).
- ⁶³ *Ibid.*
- ⁶⁴ *Ibid.*, 635, [second emphasis added].

Little Things Matter a Lot: The Significance of Implicit Bias, Practically & Legally

⁶⁵ Doing the same for Connecticut's death penalty, see also *Connecticut v. Santiago*, 318 Conn. 1 (2015).

⁶⁶ *Comcast Corporation v. National Association of African American-Owned Media, et al.*, 140 S.Ct. 1009 (2020), 1014.

⁶⁷ For thoughtful analysis, see Katie Eyer, "The But-For Theory of Antidiscrimination Law," *Virginia Law Review* 107 (8) (2021): 1621.