

Disrupting the Effects of Implicit Bias: The Case of Discretion & Policing

Jack Glaser

Police departments tend to address operational challenges with training approaches, and implicit bias in policing is no exception. However, psychological scientists have found that implicit biases are very difficult to reduce in any lasting, meaningful way. Because they are difficult to change, and nearly impossible for the decision-maker to recognize, training to raise awareness or teach corrective strategies is unlikely to succeed. Recent empirical assessments of implicit bias trainings have shown, at best, no effect on racial disparities in officers' actions in the field. In the absence of effective training, a promising near-term approach for reducing racial disparities in policing is to reduce the frequency of actions most vulnerable to the influence of bias. Specifically, actions that allow relatively high discretion are most likely to be subject to bias-driven errors. Several cases across different policing domains reveal that when discretion is constrained in stop-and-search decisions, the impact of racial bias on searches markedly declines.

For anyone considering the topic of racial bias in policing, the murder of George Floyd, a Black man, by a White police officer in Minneapolis in 2020 looms large. The killing was slow (a nine-minute strangulation) and conducted in broad daylight. There were passionate, contemporaneous pleas from the victim and onlookers. One has to wonder if any amount of antibias training could have prevented that officer from killing Mr. Floyd. In contrast is the 2018 killing of Stephon Clark in his family's backyard in Sacramento. Clearly a wrongful killing by the police, the circumstances nevertheless differ considerably from the Floyd case. It was nighttime, and Clark, a twenty-two-year-old Black man, was shot to death by police officers who rushed around a blind corner, opening fire when they putatively mistook the phone in his hand for a gun.

As jarring as these accounts are, they are only two examples of a much larger problem revealed in the aggregate statistics. Prior to 2014 – the year a police officer fatally shot Michael Brown, an unarmed eighteen-year-old Black boy, in Ferguson, Missouri, and the widespread attention the subsequent protests garnered – data on fatal incidents of police use of force were sorely inadequate. While official statistics tended to put the count of fatal officer-involved shootings at roughly five hundred

per year in the United States, a thorough accounting by *The Washington Post* (corroborated by other organizations, like Fatal Encounters) has found that the actual number is roughly double that.¹ The racial disparities in these fatal events are marked. In a typical year, victims of these shootings are disproportionately Black, and the disparity is even greater among victims who were unarmed at the time of shooting.² Policy researcher Amanda Charbonneau and colleagues reported that, among off-duty police officers who were fatally shot by on-duty officers over a period studied, eight of ten were Black, a disproportion that we estimated had a less than one-in-a-million probability of occurring by chance.³ Sociologists Frank Edwards, Hedwig Lee, and Michael Esposito used national statistics from 2013 to 2018 to estimate that the lifetime risk of being killed by police is about one in one thousand for Black men; twice the likelihood of American men overall.⁴

Fatal cases are just the tip of the iceberg. For nonfatal incidents, multiple research groups using heterogeneous methods have consistently found Black Americans to be disproportionately subject to all nonfatal levels of use of force by police.⁵

It is illuminating to further contrast the use of force and killings by police of unarmed Black men with what is, on its face, a more innocuous kind of police-civilian encounter, but one that happens with far greater frequency and has devastating cumulative effects on communities of color. These are discretionary investigative contacts, such as pedestrian and vehicle stops, many of which are based on vague pretexts like minor equipment violations or “furtive movements” that serve primarily to facilitate investigatory pat-downs or searches, most of which prove to be fruitless.⁶ This essay considers the broad range of police-civilian encounters, from the routine to the deadly, because the implications for the role of implicit bias, and the promise of the available countermeasures, vary dramatically across the spectrum.

Implicit bias trainings are unlikely to make a difference for officers who will commit murder in cold blood. But for officers who are entering a fraught use-of-force situation (or, for that matter, are faced with the opportunity to prevent or de-escalate one), having a heightened awareness about the potential for bias-driven errors, and/or having an attenuated race-crime mental association, could make the difference in a consequential split-second decision. For officers engaged in more day-to-day policing, effective interventions might help them to focus their attention on operationally, ethically, and constitutionally valid indicators of criminal suspicion and opportunities to promote public safety.

Implicit bias is real, it is pervasive, and it matters. Implicit bias (also known as automatic bias or unconscious bias) refers to mental associations between social groups (such as races, genders) and characteristics (such as good/bad, aggressive) that are stored in memory outside of conscious awareness and are activated automatically and consequently skew judgments and affect behaviors of

individuals.⁷ Other essays in this volume go into greater depth and breadth on the science and theory behind the concept of implicit bias, but I will provide here a succinct description that highlights the themes most important to efforts to disrupt implicit bias.⁸

The theoretical origins of implicit bias, a construct developed and widely used by social psychologists, are firmly planted in the sibling subfield of cognitive psychology. Cognitive psychologists interested in how people perceive, attend to, process, encode, store, and retrieve information used ingenious experimental methods to demonstrate that much of this information processing occurs outside of conscious awareness (implicitly) or control (automatically), enabling people to unknowingly, spontaneously, and effortlessly manage the voluminous flow of stimuli constantly passing through our senses.⁹

Beginning in the 1980s, social psychologists applied these theories and methods to understand how people process information about others, and in particular, with respect to the groups (racial, ethnic, gender, and so on) to which they belong.¹⁰ This research area of implicit social cognition proved tremendously effective for demonstrating that people had mental associations about social categories (such as racial groups) that could be activated automatically, even if the holder of these associations consciously repudiated them. These associations could reflect stereotypes (associations between groups and traits or behavioral tendencies) or attitudes (associations between groups and negative or positive evaluation; that is, “prejudice”).

A major advantage for the social science of intergroup bias provided by measures of implicit bias was that these methods could assess biases at a time when it was taboo to express them explicitly. At least as important, these methods measure biases people may not even know they hold and are unlikely to subjectively experience their activation or application, let alone effectively inhibit.

The methods for measuring implicit associations are indirect. In contrast to traditional methods for measuring beliefs and attitudes that involve asking people directly, or even subtler questionnaire approaches like the Modern Racism Scale, measures of implicit associations involve making inferences about the strength of the association.¹¹ That inference is based on the facility with which people process stimuli related to different categories, typically measured by the speed with which they respond to words or pictures that represent groups of people when they are paired with stimuli representing the category about which their association is being assessed.

In 1998, psychologists Anthony G. Greenwald, Debbie E. McGhee, and Jordan L. K. Schwartz published the first of very many reports of the Implicit Association Test (IAT).¹² The IAT is noteworthy because it is far and away the most widely utilized tool to assess implicit bias, and has benefited from the thorough exploration of its psychometric properties that has resulted. As described in detail by

Kate A. Ratliff and Colin Tucker Smith in this volume, the IAT yields a bias score that reflects the standardized average speed with which the participant responds when the categories are combined one way (for example, Black associated with good, White associated with bad) versus the other, thereby allowing for an inference that the individual associates one group with one trait (good or bad) more than the other.

Considering that the IAT is generating an index of the strength of someone's mental associations between categories based on the speed to press buttons in response to a disparate array of stimuli that are, by the way, presented in a different order for each participant, we do not expect it to be a *strong* predictor of anything; in scientific terms, it is "noisy," and should not be used for "diagnostic" purposes at the individual level. Nevertheless, when looking at aggregate data, the IAT and similar measures have been shown to have reasonably good construct validity and test-retest reliability.¹³

The IAT has become so influential, in part, because it has now been carried out literally millions of times through the Project Implicit website, which hosts numerous versions of the IAT that can be taken for demonstration or research purposes.¹⁴ As a result, researchers have been able to test the convergent validity of the IAT, finding that it correlates reliably and predictably with explicit (that is, direct, questionnaire-based) measures of the same attitudes.¹⁵ Therefore, although implicit bias scores are indirect, representing response speed differences to varied series of stimuli, they correlate with measures that, although subject to self-presentation bias, are clear on their face about what they are measuring.

More important than correspondence with explicit measures, which have their own limitations, implicit measures have been shown to correlate with behavior, specifically, discriminatory behavior.¹⁶ Psychologist Benedek Kurdi and colleagues carried out a meta-analysis of over two hundred studies with tests of IAT-behavior relations, finding small but consistent positive relations, above and beyond (that is, after statistically controlling for) explicit measures of bias.¹⁷ They also found that the more methodologically rigorous the study, the larger the relationship. Although these effects tend to be small, psychologists Anthony G. Greenwald, Mahzarin R. Banaji, and Brian A. Nosek, as well as legal scholar Jerry Kang in this issue, have rightly noted that small effects, when widespread and persistent, can have cumulatively large consequences.¹⁸

Some of these research findings involve correlations between implicit bias measures and highly important, real world discriminatory behaviors.¹⁹ For example, economist Dan-Olof Rooth found that implicit preference for ethnically Swedish men over Arab-Muslim men in Sweden predicted the rate at which real firms invited applicants for interviews as a function of the ethnicity conveyed by the names on otherwise identical résumés: recruiters with stronger anti-Arab-Muslim implicit bias were less likely to invite applicants with Arab-Muslim

sounding names.²⁰ Implicit racial attitudes significantly predicted self-reported vote choice in the 2008 U.S. presidential election, even after controlling for common vote predictors like party identification, ideology, and race.²¹ Implicit associations between the self and death/suicide predicted future suicide attempts in a psychiatric population.²² In a sample of medical residents, implicit racial bias was associated with a decreased tendency to recommend an appropriate treatment for a Black patient, and an increased tendency for a White patient.²³

In my own research, we have found that an implicit association between Black people and weapons (but not the generic Black-bad association) is a predictor of “shooter bias”: the tendency to select a shoot (instead of a don’t shoot) response when presented with an image of an armed Black man, as opposed to an armed White man.²⁴ Our study used a college undergraduate sample, but other studies of shooter bias have found it to be prevalent in police samples.²⁵

Another line of research has found evidence of police officers taking longer to shoot Black individuals than White individuals, and being less likely to shoot unarmed Black people than unarmed White people.²⁶ In this simulation, officers were presented with video vignettes that lasted roughly forty seconds. In each vignette, the suspect appears early, but the decision to shoot, prompted by the appearance of a weapon, for example, occurs late. Under these conditions, officers may have time to marshal corrective strategies. In contrast, the shooter bias studies involve a series of rapid responses, with each trial taking less than one second. Interestingly, the same researchers who observed the reverse-racism effect in the more protracted simulation have found that police officers associate Black people and weapons, and that the association is most pronounced when they have had relatively little sleep.²⁷ Taken together, these sets of findings suggest that at least some officers would override their implicit racial bias if given the opportunity. This is consistent with the MODE (Motivation and Opportunity as Determinants) model of information processing used to explain attitude-behavior relationships.²⁸

However, there is a well-established tradeoff between speed and accuracy when people make decisions.²⁹ Under realistic conditions, wherein there are distractions, distress, and a sense of threat, processing difficulties reflected in response latency are likely to translate into errors.

Be it in hiring, health care, voting, policing, or other consequential decision-making, implicit biases have been shown to be influential, implicating the need for effective interventions to promote nondiscrimination.

As cognitive psychologists demonstrated decades ago, implicit cognition is a constant fact of life. It serves an adaptive function of helping people manage a volume of information that would be impossible to handle consciously. It also helps us automatize the activation of memories and processes, such as driving a car, to free up conscious resources for more novel and com-

plex decisions.³⁰ This is true for memories of people and the categories in which we perceive them as belonging. As a consequence, implicit stereotypes and attitudes are pervasive. There is an extensive social psychological literature on what the sources and causes of these biases are, and there is a clear accounting of the extent of implicit bias from research using many thousands of IAT results gathered through Project Implicit.³¹

Directly relevant to the issue of implicit bias and policing, psychologists Eric Hehman, Jessica K. Flake, and Jimmy Calanchini have shown that regional variation in implicit racial bias (based on Project Implicit data) is associated with variation in racial disparities in police use of force, and psychologists Marleen Stelter, Iniobong Essien, Carsten Sander, and Juliane Degner have shown that county-level variation in both implicit and explicit prejudice is related to racial disparities in traffic stops.³² The greater the average anti-Black prejudice, the greater the ratio of stops of Black people relative to their local population. These findings do not speak conclusively to whether there is a direct, causal link between police officers' implicit bias levels and their racially disparate treatment of community members. But they suggest that, at the very least, variation in the cultural milieu that gives rise to implicit biases affects police performance as well.

Given its prevalence and influence over important behaviors, there has long been interest in identifying conditions and methods for changing implicit biases. Cognitive social psychologists have been skeptical about prospects for meaningfully and lastingly changing implicit biases because of their very nature: they reflect well-learned associations that reside and are activated outside of our subjective experience and control. Furthermore, they would not serve their simplifying function well if they were highly subject to change. Being products of what we have encountered in our environments, implicit biases are unlikely to change without sustained shifts in the stimuli we regularly encounter. For that matter, even explicit attitudes and beliefs are difficult to modify.³³ Nevertheless, considerable exploration has been conducted of the conditions under which implicit biases can change, or at least fluctuate.

One important strain of research is on the malleability of implicit biases. Distinguishable from lasting change, malleability refers to contextual and strategic influences that can temporarily alter the manifestation of implicit biases, and considerable evidence has shown that the activation and application of implicit biases are far from inevitable. For example, social psychologist Nilanjana Dasgupta has found over a series of studies that scores on measures of implicit bias can be reduced (although rarely neutralized) by exposing people to positive examples or media representations from the disadvantaged group.³⁴ Social psychologist Irene V. Blair provided an early and compelling review of implicit bias malleability, noting that studies showed variation in implicit bias scores as a func-

tion of experimenter race and positive mental imagery, and weaker implicit stereotypes after extended stereotype negation training (that is, literally saying “no” to stereotype-consistent stimulus pairings).³⁵ On the other hand, there is research showing that implicit biases are highly resistant to change.³⁶ Recent efforts to examine the conditions under which implicit attitudes may or may not shift have revealed, for example, that evaluative statements are more impactful than repeated counter-attitudinal pairings, and that change is easier to achieve when associations are novel (in other words, learned in the lab) as opposed to preexisting.³⁷

For my part, I have been interested in the possibility that egalitarian motivations can themselves operate implicitly, holding promise for automatic moderation of implicit bias effects.³⁸ Research has shown that goals and motives, like beliefs and attitudes, can operate outside of conscious awareness or control.³⁹ Furthermore, research on explicit prejudice has shown that motivation to control prejudiced responding, as measured with questionnaires, moderates the relation between implicit bias and expressed bias.⁴⁰ My colleagues and I developed a reaction time–based method to identify those who are most likely to be *implicitly* motivated to control prejudice (IMCP), finding that those who had a relatively strong implicit association between prejudice and badness (an implicit negative attitude toward prejudice) as well as a relatively strong association between themselves and prejudice (an implicit belief oneself is prejudiced) showed the weakest association between an implicit race-weapons stereotype and shooter bias.⁴¹ We further found that only those high in our measure of IMCP were able to modulate their shooter bias when their cognitive resources were depleted, providing evidence that the motivation to control prejudice can be automatized (that is, operate largely independently of cognitive resources).⁴²

Several robust efforts have been made to test for effective methods to lastingly reduce implicit bias. Social psychologists Patricia G. Devine, Patrick S. Forscher, Anthony J. Austin, and William T. L. Cox tested a multifaceted, long-duration program to “break the prejudice habit.”⁴³ They developed an approach emphasizing the importance of people recognizing bias (awareness), being concerned about it (motivation), and having specific strategies for addressing it. Their program took place over an eight-week span as part of an undergraduate course, and they found significant reductions in (albeit, by no means elimination of) implicit bias four and eight weeks after the beginning of the program. However, a subsequent intervention experiment on gender bias among university faculty, while still showing promising effects on explicit and behavioral measures, did not replicate reductions in implicit bias.⁴⁴

With respect to focused, short-term methods for reducing implicit bias, some extraordinarily systematic research has been conducted, finding that some approaches can partially reduce implicit racial bias, but that these effects are fleeting.⁴⁵ Social psychologist Calvin K. Lai and colleagues coordinated a “many labs”

collaboration to test a set of seventeen promising strategies to reduce implicit bias, specifically, the Black/White–bad/good association. The strategies include multiple methods to help participants engage with others’ perspectives, expose them to counter-stereotypical examples, appeal to egalitarian values, recondition their evaluative associations, induce positive emotions, or provide ways to override biases. Additionally, an eighteenth strategy, “faking” the IAT, was tested. At least three research groups tested each strategy, allowing for statistically powerful, reliable inferences. While nine of these eighteen approaches yielded virtually no change in implicit bias as measured on the IAT, the other nine yielded statistically significant, albeit only partial, reductions. However, in a subsequent, careful, and robust study, Lai and colleagues retested the nine effective strategies, finding, first, that all were again able to cause statistically significant reductions in implicit bias, but that when the IAT was administered between two and twenty-four hours after the initial test, all but one of the groups’ implicit bias scores had returned to baseline – the bias reduction effects were partial and short-lived.⁴⁶ Similarly, social psychologist Patrick S. Forscher and colleagues conducted a large meta-analysis of experiments testing methods to reduce scores on implicit bias measures, finding the typical effects to be weak.⁴⁷

This is not by any means conclusive evidence that bias reduction strategies cannot have substantial, lasting effects, perhaps with the right dosing (duration and repetition). However, the body of evidence to date indicates that, without meaningful, lasting environmental change, implicit biases are resilient. This is entirely consistent with the theory and evidence regarding implicit cognition more generally: the ability to store, activate, and apply implicit memories automatically is adaptive. If implicit associations, particularly those well-learned (such as over a significant period of time), were highly malleable or changeable, they would not serve their function.

In policing, as in many other industries, providing trainings is a method of first resort when concerns about discrimination arise. Unfortunately, few of these trainings are accompanied by rigorous evaluations, let alone assessments including behavioral or performance outcomes.⁴⁸ Some systematic reviews of diversity trainings have found small effects on behavioral outcomes. Psychologist Zachary T. Kalinoski and colleagues found small- to medium-sized effects for “on-the-job behavior” in the six studies in their meta-analysis that included such behavioral outcomes.⁴⁹ In a large meta-analysis of diversity training program studies, psychologist Katerina Bezrukova and colleagues found relatively small effects on behavioral outcomes.⁵⁰ On the other hand, in their large-scale study, sociologists Alexandra Kalev, Frank Dobbin, and Erin Kelly found that diversity training had no effect on the racial or gender managerial composition of firms.⁵¹ Psychologist Elizabeth Levy Paluck and colleagues have carefully reviewed the effects of diver-

sity trainings, finding few to have meaningful measures of behavioral outcomes, and for those few to be lacking evidence of effects on actual behavior.⁵²

In policing, there has been considerable participation in diversity training, with much of it labeled as “implicit bias training,” in particular. CBS News surveyed a sample of one hundred fifty-five large American municipal police departments, finding that 69 percent reported having carried out implicit bias trainings.⁵³ Departments and trainers, however, have not participated in robust evaluations of the effects of implicit bias training on officer performance, until recently. In 2018–2019, under the supervision of a court-appointed monitor resulting from a civil suit, one of the world’s largest law enforcement agencies, the New York Police Department (NYPD), engaged the industry leader Fair and Impartial Policing in implicit bias training for its roughly thirty-six thousand sworn officers.⁵⁴ The effects of the training were evaluated effectively by exploiting the staggered rollout of the program, allowing for a comparison of field performance for officers before and after the training without confounding the comparison with any particular events that occurred simultaneously.⁵⁵ The researchers found that, while officers evaluated the training positively and reported greater understanding of the nature of implicit bias, only 27 percent reported attempting to apply their new training frequently (31 percent “sometimes”) in the month following, while 42 percent reported not at all. More concerning, comparisons between pre- and post-training of the racial distributions of those stopped, frisked, searched, and who had force used against them revealed that, if anything, the percent who were Black increased. This study occurred from 2017 to 2019, a period after which the controversial stop-question-and-frisk (SQF) program had been ruled unconstitutional and dramatically reigned in, so disparities had already been somewhat reduced, leaving less room for improvement. However, as the study data reveal, while Black people – who make up about 25 percent of the city population – were 59.3 percent of those stopped in the first six months of 2019, they were only 47.6 percent of those arrested, suggesting that there remained considerable racial bias in who was being stopped.

A very recent, rigorous evaluation of the effects of another mainstream implicit bias training for police was conducted by Calvin K. Lai and Jaclyn A. Lisnek.⁵⁶ In this study, while trained officers indicated greater knowledge of bias that lasted at least one month after training, their increased concerns about bias, and understanding of the durability of bias, were more fleeting. With respect to behavioral outcomes, while officers indicated intentions to use strategies to manage bias following the training, their self-reported actual use of the strategies in the month after training was, disappointingly, lower than their self-reported use at baseline (prior to training).

The null effects on behavior come as no surprise to cognitive social psychologists, given that these trainings typically aim to, in a single day or less, mitigate the effects of cognitive biases that are learned over the lifespan, operate outside of

conscious awareness, and occur automatically. That said, implicit bias is not the only cause of discrimination, so it is especially discouraging that these trainings, which emphasize the importance of bias and awareness of it, do not appear to affect behavior through other channels, such as conscious, deliberate thought and behavior. On the other hand, this should not be all that surprising, given the very subtle and mixed effects of other forms of prejudice reduction trainings. This is not to say that implicit bias and other prejudice reduction trainings have no hope of meaningfully and lastingly reducing discrimination. There will need to be, however, further development and testing of training strategies that work. Until then, other avenues for disrupting implicit bias must also be explored.

In the absence of training that meaningfully and, ideally, lastingly reduces disparate treatment, a promising approach to reduce the impact of implicit biases is to constrain discretion. As Amanda Charbonneau and I, and others, have explained, police officers have a high degree of discretion (that is, latitude) in how they conduct their duties.⁵⁷ This stems in part from the vagueness of the regulatory standards, particularly “reasonable suspicion,” that govern their practices. Suspicion is an inherently subjective experience, and its modifier “reasonable” is an intentionally vague standard that is often tautologically defined: “reasonable” is what a reasonable person (or officer) would think or do. Many people, in their professional endeavors, have discretion in how they carry out their jobs, including decisions about academic grading, admissions, and hiring; public- and private-sector hiring and promotions; legislative voting; public benefits eligibility; and mental and physical health care. Although individual professionals gain expertise through training and experience that may help them make good assessments, we rarely make decisions with complete information, and the evidence is clear that, in the absence of complete and specific information, we often rely on cognitive shortcuts like stereotypes, and/or interpret evidence in ways that are consistent with our prior conceptions or preferred outcomes.⁵⁸

When discretion is high – for example, when decision-makers can use their own judgment in ambiguous situations – cognitive shortcuts like stereotypes have more opportunity to influence decisions. Analyses of real-world data on hiring and disciplinary decisions demonstrate that, in the absence of specific information, biases are influential. For example, economists Harry J. Holzer, Steven Raphael, and Michael A. Stoll found that employers who carried out criminal background checks were *more* likely to hire African Americans, suggesting that, in the absence of specific information about criminality, decision-makers may make the stereotype-consistent assumption.⁵⁹ With the specific information, they are less likely to discriminate. Similarly, economist Abigail Wozniak found that the implementation of legislation promoting drug testing resulted in substantial increases in Black employment rates, again raising the possibility that, in the absence of specific infor-

mation, stereotype-consistent judgments will disadvantage stigmatized groups in high-discretion decision-making like hiring, promotion, and retention.⁶⁰

In the domain of school discipline, which bears important similarities and even a direct relationship to criminal justice (that is, the school-to-prison pipeline), psychologist Erik J. Girvan and colleagues Cody Gion, Kent McIntosh, and Keith Smolkowski found that, in a large dataset of school discipline cases, the vast majority of the variance in racial disparities was captured in high-discretion referrals.⁶¹ Specifically, cases involving indicators of misconduct that were determined by the subjective assessment of school staff, as opposed to those with objective criteria, had far more racially disparate referral rates.

Specific to policing, Charbonneau and I have considered three large cases in which officer discretion can be operationalized in different ways.⁶² We found that, across a range of law enforcement agencies, higher discretion in decisions to search was associated with greater disparities in search yield rates. Specifically, when discretion was high, White people who were searched were more likely to be found with contraband than were Black people or Latino people. In two of these cases (U.S. Customs and New York City), policy changes allow for a reasonably strong causal inference that reductions in discretion reduce disparities.

Comparisons of search yield rates (the percentage of searches that yield contraband) offer a compelling method to identify bias in law enforcement decisions. Drawing from the larger research literature on “outcome tests,” the inference can be made that, if searches of one group of people are more likely to result in findings of contraband, then whatever is giving rise to decisions to search members of that group is generally a better indicator of criminal suspicion than whatever is triggering searches of other groups.⁶³ In other words, groups with higher search yield rates are probably being subjected to higher thresholds of suspicion in order to be searched. Groups who are searched based on lower levels of individual suspiciousness (perhaps because their group is stereotyped as prone to crime) will be less likely to be found in possession of evidence of crime. In turn, if one group has lower search yield rates than others, it can be inferred that there is group-based bias in at least some of the decisions to search. This could be compounded by group-based bias in decisions to surveil and stop, in the first place. When high discretion of who to surveil, stop, and search is afforded to officers, these decisions will be made under higher degrees of ambiguity (that is, less determined by codified criteria) and will therefore be more prone to the influence of biases such as racial stereotypes, thereby causing disparities.

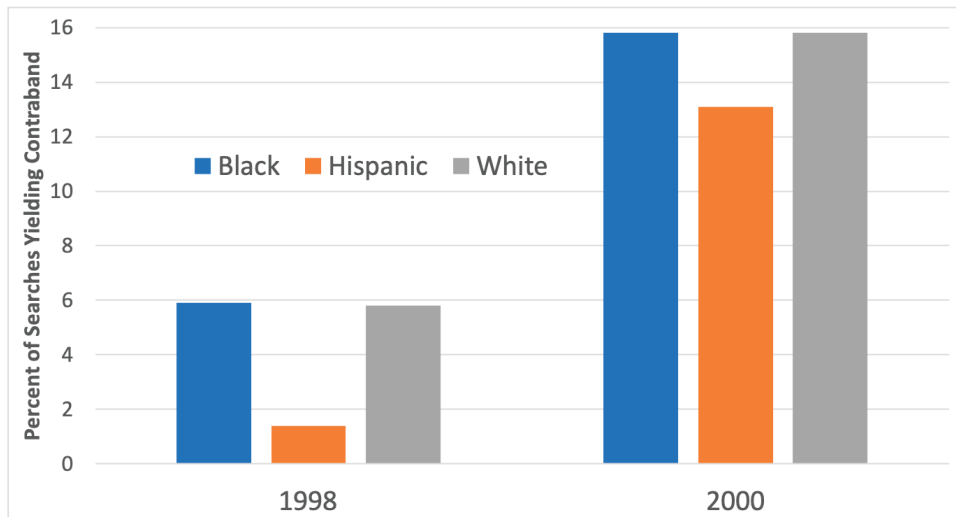
In 1999, the U.S. Customs Service (now Customs and Border Patrol) reduced the number of criteria for triggering a search of a traveler from forty-three to six, with the new criteria being more instrumentally related to smuggling.⁶⁴ Comparing the full year before to the full year after the reduction in search criteria, the number of searches declined 75 percent, but the search yield rate quadru-

pled. More important, the search yield rates became much less racially disparate. Prior to the change, the search yield rates for Hispanic people had been roughly one-quarter of the rate for Black people and White people who were searched, strongly suggesting that Hispanic people were being searched at lower thresholds of suspiciousness (because their searches were less likely to prove to be justified). As shown in Figure 1, after the reduction in search criteria, search yield rates increased overall, and nearly equalized across groups. Part of the inequity may have been due to a large share of Customs searches occurring at the U.S.-Mexico border, where searches may have been more frequent overall. However, the dramatic reduction in search yield disparities after the change in search criteria indicates that the disparity was mostly due to differential standards of suspicion being applied when discretion was high – when there were a lot of criteria to choose from. If the disparity had been due solely to different rates of searches at different ports of entry, the change in criteria would not have caused nearly as large a reduction in yield rate disparities.

The effects in the U.S. Customs case, in terms of increased yields overall and decreased disparities, are dramatic. This may be due in part to the nature of customs searches, which involve a decision (to search or not) about each person passing through the system, in contrast to searches in traffic enforcement or street policing, in which the decision to search is conditional upon the decisions to surveil and stop, which are also based on suspicion. In the noisier latter condition, the effects of discretion on search yield disparities would likely be smaller.

The second case involves the largest law enforcement agency in the United States. Over several decades, the NYPD has had an ebbing and flowing of the stop-question-and-frisk program, involving thousands of low-level stops of pedestrians with the primary goal of reducing street crime. A wave of increasing SQF began in the early 2000s, peaking in 2011 with over 685,000 stops in one year. About half of those stopped were Black people (mostly young men) – double their rate of residency – and about half of all of those stopped were subjected to frisks or searches, but the rate was considerably higher for Black people and Latino people than for White people. Officers most often recorded using highly subjective criteria, such as furtive movements, to justify their stops. Due to shifting political winds and a successful class action lawsuit, SQF declined (at least as indicated by reported stops) precipitously after 2011, plummeting to fewer than 20,000 stops per year by 2015.⁶⁵ As is commonly the case in search yield statistics, contraband and weapon discovery rates in 2011 were much higher for White people who were frisked than for Black people or Latino people. The White-Black yield rate ratio was 1.4-to-1 and 1.8-to-1 for contraband and weapons, respectively. In 2015, with a fraction of the number of stops, and the removal of furtive movements and other high-discretion reportable bases for stops, those ratios declined to 1.1-to-1 and 1.05-to-1, indicating that decisions to stop and frisk were less influenced by racial bias.

Figure 1
Percent of U.S. Customs Searches Yielding Contraband Before and After the 1999 Reduction in Search Criteria



Source: Author's image, based on data from Deborah Ramirez, Jennifer Hoopes, and Tara Lai Quinlan, "Defining Racial Profiling in a Post-September 11 World," *American Criminal Law Review* 40 (2003): 1195.

Especially telling are our analyses of statewide data from California, facilitated by the 2015 passage of the Racial and Identity Profiling Act (RIPA) requiring all law enforcement agencies in the state to report data on all traffic and pedestrian stops.⁶⁶ In contrast to the U.S. Customs and NYPD cases, where we compared racial disparities in search yield rates as a function of reduced discretion in search practices over time, with the RIPA data, we compared disparities across search types that varied in how discretionary they tend to be. For example, reviewing data from the first wave of RIPA – the eight largest departments in the state (including the Los Angeles Police Department, LA County Sheriff, and California Highway Patrol) – we found that yield rates were higher for White people than Black people and Latino people for searches based on supervision status (such as probation or parole), which allow officers considerable discretion, first to ask if someone is under supervision, and then to opt to search. However, for searches that were “procedural,” such as those required during an arrest (“incident to arrest”), the search yield rates for White people were comparable to those for Black people and Latino people.⁶⁷

Across these three cases, including a large, federal agency, an immense metropolitan police department, and the eight largest agencies in the most populous American state, we found that when officers' search discretion was relatively high, White people who were searched were more likely to be found in possession of contraband or weapons, indicating that White people were being subjected to higher thresholds of suspicion than Black people and Latino people in order to get stopped and/or searched. When discretion was relatively low (when search decisions were based on more stringent, prescribed criteria), yield rates were higher overall, and far less disparate. The evidence reviewed indicates that reducing discretion – in police stop-and-search practices, school discipline, private-sector hiring, and likely many other domains – is an effective method for reducing racial, ethnic, or other disparities. In the policing cases, at least, the overall improvements in search yield rates when discretion is low suggest that the effectiveness of the work need not be compromised. This was literally the case in Customs searches because, while searches dropped 75 percent, contraband discoveries quadrupled, resulting in roughly the same raw number of discoveries. That reductions in searches will have commensurate increases in yields is by no means likely, let alone guaranteed. This was certainly not the case in New York City, where the roughly 97 percent decline in pedestrian stops was accompanied by approximately a doubling in search yield rates. However, concerns that reducing SQF would result in an increase in crime were not borne out.⁶⁸ In fact, the continued decline in crime following SQF's near elimination was compelling enough to cause some rare public *mea culpas*.⁶⁹ It should also be noted that a large majority of the contraband recovered in NYPD searches was drug-related, while firearm seizures numbered in the hundreds, even at the peak of SQF. Even if high-discretion searches have, under some circumstances, a deterrent effect on crime, this must be weighed against the psychological harms caused by overpolicing, not to mention the violations of Fourth and Fourteenth Amendment protections against unreasonable searches and seizures and of equal protection.

When considering what can and cannot be done to disrupt the effects of implicit biases, it is crucial to bear in mind that implicit biases cause discriminatory judgments and actions *indirectly*. Because they operate outside of conscious awareness and control, and are generally not subjectively experienced by their holders, their effects are largely unintentional. Even an overt racist can have his bigotry enhanced (or possibly diminished) by implicit biases of which he is not aware.

An illustrative example of how implicit bias causes discrimination comes from a classic experiment that preceded the implicit bias innovations in psychological science. Psychologists John M. Darley and Paget H. Gross had research subjects evaluate the academic performance of a schoolgirl ostensibly named Hannah. Half of the sample was led to believe Hannah was from a low socioeconomic status

(SES) background, and the other half from a high SES background.⁷⁰ Splitting the sample yet again, half in each SES condition gave estimates of how they thought Hannah would do, while the other half rated her performance after watching a video of Hannah taking the tests. Among those who predicted Hannah's performance without watching the video, the low and high SES groups rated her about the same. Among those who actually observed her performance, even though all research participants watched the identical video, those who were given the impression that Hannah was low SES tended to rate her performance as below grade level, and those who were led to think she was high SES tended to rate her performance above grade level. They watched the same video, but interpreted the ambiguities in her performance in ways consistent with their stereotypes of low and high SES children. This was not intentional, or there would have been a similar pattern for those who did not see the video. People were, probably in good faith, doing their best to appraise Hannah's performance given the information they had. Their information about her socioeconomic status and the associated stereotypes skewed their perceptions. Likewise, implicit biases we may not even know we have, let alone endorse, can skew our perceptions and cause discriminatory judgments and behaviors.

This reality helps to explain how company hiring managers and staff will be inclined to interview people who have White- as opposed to Black-sounding names despite their résumés being identical, and why employers might tend to assume that Black applicants have criminal backgrounds or are drug users.⁷¹ In the case of policing, officers are more likely to assume that people of color are involved in crime, even though searches of these individuals rarely bear this out, and they typically yield more evidence of criminality among White people who are searched – because the searches are biased.

In the absence of reliable methods for eliminating implicit (or, for that matter, explicit) biases, and with research indicating that trainings promoting cultural awareness, diversity, and fairness do not reliably reduce disparities in the real world, minimizing the vulnerability factors for discrimination is the best option. Reducing discretion and, ideally, replacing it with prescriptive guidance and systematic information (that is, valid criteria) has been shown to be effective with respect to stop-and-search decisions in policing.

Use of lethal force may require a special variant on the approach of reducing discretion. As discussed above with respect to demonstrations of police officers exhibiting “shooter bias” in simulations, situations in which police use force, and especially those that may involve lethal force, are fraught with vulnerabilities to errors. These situations typically involve time pressure, distraction, cognitive load, and intense emotions, including fear and anger. Many of these situations occur at night, adding visual ambiguity and heightening uncertainty and fear. As Jennifer T. Kubota describes, for many White Americans,

mere exposure to the image of a Black person's face triggers neurological activity consistent with fear, and the differential fear response to Black faces compared to White faces or neutral objects has been found to be associated with implicit racial bias.⁷² This automatic fear response, occurring even in a mundane laboratory setting, is surely compounded by the anticipated (and often exaggerated) sense of mortal threat that police bring to civilian encounters.⁷³

Given that implicit bias trainings for police, or even officers' self-reported utilization of trained strategies to interrupt bias, have been shown not to reduce disparate outcomes in stop, search, arrest, and use of nonlethal force, limiting the discretion with which police officers use force needs to be prioritized. In California, state law has been changed to require that lethal force be employed only when "necessary," a more stringent criterion than what it replaced: "reasonable."⁷⁴ However, it remains to be seen if this statutory change will translate into reduced levels of, and disparities in, excessive force, or if courts will merely apply a reasonableness standard to the necessity criterion (like what a "reasonable" officer would deem "necessary").

Some police departments appear to have had success in developing intensive trainings that reduce the unnecessary use of lethal force.⁷⁵ Practitioners emphasize the importance of officers slowing things down, keeping distance, and finding cover to reduce the likelihood of unnecessary force being used – approaches reflective of the challenges that the automatic activation of implicit racial bias presents. De-escalation training is also popular. But while there is at least one example of a police training program demonstrated to have reduced use of force and its collateral consequences, such as injuries, the evidence of trainings' effectiveness in general has been unclear.⁷⁶ To the extent that use of force is applied in a racially disparate manner, and the evidence of that is clear, reductions in unnecessary force should reduce disparities, just as reductions in unnecessary searches do.⁷⁷ Even if implicit bias is a substantial cause of disparities in police officers' use of force, interventions that directly target implicit bias are unlikely to succeed.

Researchers and practitioners can, and will, keep trying to look for practicable ways to reduce and/or override implicit biases through training. Some have made inroads, although the long-term effects on implicit biases themselves are tenuous, at best. While we wait for breakthroughs in methods and dosing, identifying institutional and personal vulnerabilities (such as hiring practices, enforcement practices, incentives, habits, distractions, cognitive load, and decision points) and possible methods to address them (for example, through constraints on discretion and prescriptions for better approaches) is more promising given the current state of the field. Rebecca C. Hetey, MarYam G. Hamedani, Hazel Rose Markus, and Jennifer L. Eberhardt describe prime examples of these kinds of prescriptive interventions in their contribution to this volume, including

requiring that officers provide more extensive explanations for their investigative stops.⁷⁸ As Hetey and coauthors as well as Manuel J. Galvan and B. Keith Payne argue in their essays in this issue, even if we could effectively disrupt implicit bias, we have to consider that structural factors such as historical inequities, incentives to punitiveness, and hierarchical institutional cultures are likely to be more influential than individual-level factors like implicit stereotyping. That said, individual and structural causes of discrimination are mutually reinforcing: structural inequities reinforce the negative attitudes, even at the implicit level, and vice versa.⁷⁹ Addressing structural factors can reduce considerable harm in the near future and, by attenuating disparities, possibly serve to soften individual-level biases, making them more conducive to change.

ABOUT THE AUTHOR

Jack Glaser is Professor at the Goldman School of Public Policy at the University of California, Berkeley. He is the author of *Suspect Race: Causes and Consequences of Racial Profiling* (2015), and has published in such journals as *Social and Personality Psychology Compass*, *Law & Human Behavior*, and *Policy Insights from Behavioral and Brain Sciences*.

ENDNOTES

- ¹ “Police Shootings Database 2015–2023,” *The Washington Post*, <https://www.washingtonpost.com/graphics/investigations/police-shootings-database> (accessed December 1, 2023).
- ² As Charbonneau and colleagues report in 2015, the percentage of victims who were unarmed was more than twice as high for Black people than for White people, although that disparity declined in 2016, as the overall proportion of fatal officer-involved-shootings involving unarmed victims declined. Amanda Charbonneau, Katherine Spencer, and Jack Glaser, “Understanding Racial Disparities in Police Use of Lethal Force: Lessons from Fatal Police-on-Police Shootings,” *Journal of Social Issues* 73 (4) (2017): 744–767.
- ³ Originally published in Christopher Stone, Zachary Carter, Thomas Belfiore, et al., *Reducing Inherent Danger: Report of the Task Force on Police-On-Police Shootings* (New York: New York State Task Force on Police-on-Police Shootings, 2010).
- ⁴ Frank Edwards, Hedwig Lee, and Michael Esposito, “Risk of Being Killed by Police Use of Force in the United States by Age, Race–Ethnicity, and Sex,” *Proceedings of the National Academy of Sciences* 116 (34) (2019): 16793–16798.
- ⁵ For example, Roland G. Fryer Jr., “An Empirical Analysis of Racial Differences in Police Use of Force,” *Journal of Political Economy* 127 (3) (2019): 1210–1261; Amanda Geller, Phillip Atiba Goff, Tracey Lloyd, et al., “Measuring Racial Disparities in Police Use of Force: Methods Matter,” *Journal of Quantitative Criminology* 37 (2021): 1083–1113.
- ⁶ Jack Glaser, *Suspect Race: Causes and Consequences of Racial Profiling* (Oxford: Oxford University Press, 2015).

- ⁷ Anthony G. Greenwald and Mahzarin R. Banaji, "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes," *Psychological Review* 102 (1) (1995): 4.
- ⁸ 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 and Colin Tucker Smith, "The Implicit Association Test," *Dædalus* 153 (1) (Winter 2024): 51–64, <https://www.amacad.org/publication/implicit-association-test>; and Rebecca C. Hetey, MarYam G. Hamedani, Hazel Rose Markus, and Jennifer L. Eberhardt, "'When the Cruiser Lights Come On': Using the Science of Bias & Culture to Combat Racial Disparities in Policing," *Dædalus* 153 (1) (Winter 2024): 123–150, <https://www.amacad.org/publication/when-cruiser-lights-come-using-science-bias-culture-combat-racial-disparities-policing>.
- ⁹ For example, David E. Meyer and Roger W. Schvaneveldt, "Meaning, Memory Structure, and Mental Processes: People's Rapid Reactions to Words Help Reveal How Stored Semantic Information Is Retrieved," *Science* 192 (4234) (1976): 27–33; and James H. Neely, "Semantic Priming and Retrieval from Lexical Memory: Roles of Inhibitionless Spreading Activation and Limited-Capacity Attention," *Journal of Experimental Psychology: General* 106 (3) (1977): 226.
- ¹⁰ Patricia G. Devine, "Stereotypes and Prejudice: Their Automatic and Controlled Components," *Journal of Personality and Social Psychology* 56 (1) (1989): 5; and John F. Dovidio, Nancy Evans, and Richard B. Tyler, "Racial Stereotypes: The Contents of Their Cognitive Representations," *Journal of Experimental Social Psychology* 22 (1) (1986): 22–37.
- ¹¹ John B. McConahay, Betty B. Hardee, and Valerie Batts, "Modern Racism Scale," *Personality and Social Psychology Bulletin* (1980).
- ¹² Anthony G. Greenwald, Debbie E. McGhee, and Jordan L. K. Schwartz, "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test," *Journal of Personality and Social Psychology* 74 (6) (1998): 1464–1480.
- ¹³ William A. Cunningham, Kristopher J. Preacher, and Mahzarin R. Banaji, "Implicit Attitude Measures: Consistency, Stability, and Convergent Validity," *Psychological Science* 12 (2) (2001): 163–170.
- ¹⁴ In this volume, see Morehouse and Banaji, "The Science of Implicit Race Bias."
- ¹⁵ Brian A. Nosek, "Moderators of the Relationship between Implicit and Explicit Evaluation," *Journal of Experimental Psychology: General* 134 (4) (2005): 565; Wilhelm Hofmann, Bertram Gawronski, Tobias Gschwendtner, et al., "A Meta-Analysis on the Correlation between the Implicit Association Test and Explicit Self-Report Measures," *Personality and Social Psychology Bulletin* 31 (10) (2005): 1369–1385; and 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.
- ¹⁶ For parallel with the term "explicit measures," and reflecting common usage, I will sometimes use the term "implicit measure," but recognize that these methods are more accurately described as *indirect measures of implicit bias* because in many cases, it is evident to participants what the task is measuring. The bias being measured is implicit, in the sense that it is not consciously accessible, but the measure can be downright obtrusive.

- ¹⁷ Benedek Kurdi, Allison E. Seitchik, Jordan R. Axt, et al., "Relationship between the Implicit Association Test and Intergroup Behavior: A Meta-Analysis," *American Psychologist* 74 (5) (2019): 569–586.
- ¹⁸ 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): 553–561; and Jerry Kang, "Little Things Matter a Lot: The Significance of Implicit Bias, Practically & Legally," *Daedalus* 153 (1) (Winter 2024): 193–212, <https://www.amacad.org/publication/little-things-matter-lot-significance-implicit-bias-practically-legally>.
- ¹⁹ For a review, see John T. Jost, Laurie A. Rudman, Irene V. Blair, et al., "The Existence of Implicit Bias Is Beyond Reasonable Doubt: A Refutation of Ideological and Methodological Objections and Executive Summary of Ten Studies That No Manager Should Ignore," *Research in Organizational Behavior* 29 (2009): 39–69.
- ²⁰ Dan-Olof Rooth, "Automatic Associations and Discrimination in Hiring: Real World Evidence," *Labour Economics* 17 (3) (2010): 523–534.
- ²¹ Christopher Finn and Jack Glaser, "Voter Affect and the 2008 U.S. Presidential Election: Hope and Race Mattered," *Analyses of Social Issues and Public Policy* 10 (1) (2010): 262–275.
- ²² Matthew K. Nock, Jennifer M. Park, Christine T. Finn, et al., "Measuring the Suicidal Mind: Implicit Cognition Predicts Suicidal Behavior," *Psychological Science* 21 (4) (2010): 511–517.
- ²³ Alexander R. Green, Dana R. Carney, Daniel J. Pallin, et al., "Implicit Bias among Physicians and Its Prediction of Thrombolysis Decisions for Black and White Patients," *Journal of General Internal Medicine* 22 (9) (2007): 1231–1238.
- ²⁴ Jack Glaser and Eric D. Knowles, "Implicit Motivation to Control Prejudice," *Journal of Experimental Social Psychology* 44 (1) (2008): 164–172; and Joshua Correll, Bernadette Park, Charles M. Judd, and Bernd Wittenbrink, "The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals," *Journal of Personality and Social Psychology* 83 (6) (2002): 1314–1329.
- ²⁵ Joshua Correll, Bernadette Park, Charles M. Judd, et al., "Across the Thin Blue Line: Police Officers and Racial Bias in the Decision to Shoot," *Journal of Personality and Social Psychology* 92 (6) (2007): 1006–1023; and E. Ashby Plant and B. Michelle Peruche, "The Consequences of Race for Police Officers' Responses to Criminal Suspects," *Psychological Science* 16 (3) (2005): 180–183.
- ²⁶ Lois James, Stephen M. James, and Bryan J. Vila, "The Reverse Racism Effect: Are Cops More Hesitant to Shoot Black than White Suspects?" *Criminology & Public Policy* 15 (2) (2016): 457–479.
- ²⁷ Lois James, "The Stability of Implicit Racial Bias in Police Officers," *Police Quarterly* 21 (1) (2018): 30–52.
- ²⁸ Michael A. Olson and Russell H. Fazio, "Implicit and Explicit Measures of Attitudes: The Perspective of the MODE Model," in *Attitudes: Insights from the New Implicit Measures*, ed. Richard E. Petty, Russel H. Fazio, and Pablo Briñol (New York: Psychology Press, 2008), 39–84.
- ²⁹ Robert Sessions Woodworth, "Accuracy of Voluntary Movement," *The Psychological Review: Monograph Supplements* 3 (3) (1899): i–144.

- ³⁰ John A. Bargh and Tanya L. Chartrand, "The Unbearable Automaticity of Being," *American Psychologist* 54 (7) (1999): 462–479; and Susan T. Fiske and Shelley E. Taylor, *Social Cognition* (Boston: McGraw-Hill Book Company, 1991).
- ³¹ Laurie A. Rudman, "Sources of Implicit Attitudes," *Current Directions in Psychological Science* 13 (2) (2004): 79–82; Tessa E. S. Charlesworth and Mahzarin R. Banaji, "Patterns of Implicit and Explicit Attitudes: I. Long-Term Change and Stability from 2007 to 2016," *Psychological Science* 30 (2) (2019): 174–192; 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. See also, in this volume, Morehouse and Banaji, "The Science of Implicit Race Bias."
- ³² Eric Hehman, Jessica K. Flake, and Jimmy Calanchini, "Disproportionate Use of Lethal Force in Policing Is Associated with Regional Racial Biases of Residents," *Social Psychological and Personality Science* 9 (4) (2018): 393–401; and Marleen Stelter, Iniobong Essien, Carsten Sander, and Juliane Degner, "Racial Bias in Police Traffic Stops: White Residents' County-Level Prejudice and Stereotypes Are Related to Disproportionate Stopping of Black Drivers," *Psychological Science* 33 (4) (2022): 483–496.
- ³³ Alice H. Eagly and Shelly Chaiken, "Attitude Strength, Attitude Structure, and Resistance to Change," *Attitude Strength: Antecedents and Consequences* 4 (2) (1995): 413–432.
- ³⁴ Nilanjana Dasgupta, "Implicit Attitudes and Beliefs Adapt to Situations: A Decade of Research on the Malleability of Implicit Prejudice, Stereotypes, and the Self-Concept," *Advances in Experimental Social Psychology* 47 (2013): 233–279.
- ³⁵ Irene V. Blair, "The Malleability of Automatic Stereotypes and Prejudice," *Personality and Social Psychology Review* 6 (3) (2002): 242–261; Brian S. Lowery, Curtis D. Hardin, and Stacey Sinclair, "Social Influence Effects on Automatic Racial Prejudice," *Journal of Personality and Social Psychology* 81 (5) (2001): 842; Irene V. Blair, Jennifer E. Ma, and Alison P. Lenton, "Imagining Stereotypes Away: The Moderation of Implicit Stereotypes through Mental Imagery," *Journal of Personality and Social Psychology* 81 (5) (2001): 828–841; and Kerry Kawakami, John F. Dovidio, Jasper Moll, Sander Hermsen, and Abby Russin, "Just Say No (To Stereotyping): Effects of Training in the Negation of Stereotypic Associations on Stereotype Activation," *Journal of Personality and Social Psychology* 78 (5) (2000): 871–888.
- ³⁶ Aiden P. Gregg, Beate Seibt, and Mahzarin R. Banaji, "Easier Done than Undone: Asymmetry in the Malleability of Implicit Preferences," *Journal of Personality and Social Psychology* 90 (1) (2006): 1–20; and Jennifer A. Joy-Gaba and Brian A. Nosek, "The Surprisingly Limited Malleability of Implicit Racial Evaluations," *Social Psychology* 41 (3) (2010): 137–146.
- ³⁷ Benedek Kurdi and Mahzarin R. Banaji, "Repeated Evaluative Pairings and Evaluative Statements: How Effectively Do They Shift Implicit Attitudes?" *Journal of Experimental Psychology: General* 146 (2) (2017): 194–213; and Benedek Kurdi, Kirsten N. Morehouse, and Yarrow Dunham, "How Do Explicit and Implicit Evaluations Shift? A Preregistered Meta-Analysis of the Effects of Co-Occurrence and Relational Information," *Journal of Personality and Social Psychology* 124 (6) (2022): 1174–1202.
- ³⁸ Jack Glaser and Mahzarin R. Banaji, "When Fair Is Foul and Foul Is Fair: Reverse Priming in Automatic Evaluation," *Journal of Personality and Social Psychology* 77 (4) (1999): 669–687; Jack Glaser and John F. Kihlstrom, "Compensatory Automaticity: Unconscious Volition Is Not an Oxymoron," *The New Unconscious* (2005): 171–195; and Gordon B. Moskowitz, Peter M. Gollwitzer, Wolfgang Wasel, and Bernd Schaal, "Preconscious Control of Stereotype Activation through Chronic Egalitarian Goals," *Journal of Personality and Social Psychology* 77 (1) (1999): 167–184.

- ³⁹ Tanya L. Chartrand and John A. Bargh, "Automatic Activation of Impression Formation and Memorization Goals: Nonconscious Goal Priming Reproduces Effects of Explicit Task Instructions," *Journal of Personality and Social Psychology* 71 (3) (1996): 464; and James Y. Shah and Arie W. Kruglanski, "When Opportunity Knocks: Bottom-Up Priming of Goals by Means and Its Effects on Self-Regulation," *Journal of Personality and Social Psychology* 84 (6) (2003): 1109.
- ⁴⁰ Russell H. Fazio, Joni R. Jackson, Bridget C. Dunton, and Carol J. Williams, "Variability in Automatic Activation as an Unobtrusive Measure of Racial Attitudes: A Bona Fide Pipeline?" *Journal of Personality and Social Psychology* 69 (6) (1995): 1013; and E. Ashby Plant and Patricia G. Devine, "Internal and External Motivation to Respond without Prejudice," *Journal of Personality and Social Psychology* 75 (3) (1998): 811.
- ⁴¹ Glaser and Knowles, "Implicit Motivation to Control Prejudice."
- ⁴² Sang Hee Park, Jack Glaser, and Eric D. Knowles, "Implicit Motivation to Control Prejudice Moderates the Effect of Cognitive Depletion on Unintended Discrimination," *Social Cognition* 26 (4) (2008): 401–419.
- ⁴³ Patricia G. Devine, Patrick S. Forscher, Anthony J. Austin, and William T. L. Cox, "Long-Term Reduction in Implicit Race Bias: A Prejudice Habit-Breaking Intervention," *Journal of Experimental Social Psychology* 48 (6) (2012): 1267–1278.
- ⁴⁴ Patrick S. Forscher, Chelsea Mitamura, Emily L. Dix, et al., "Breaking the Prejudice Habit: Mechanisms, Timecourse, and Longevity," *Journal of Experimental Social Psychology* 72 (2017): 133–146.
- ⁴⁵ Calvin K. Lai, Maddalena Marini, Steven A. Lehr, et al., "Reducing Implicit Racial Preferences: I. A Comparative Investigation of 17 Interventions," *Journal of Experimental Psychology: General* 143 (4) (2014): 1765.
- ⁴⁶ One of the strategies, "using implementation intentions," showed a barely significant reduced bias at the second assessment; and this effect was much reduced from the first assessment. Calvin K. Lai, Allison L. Skinner, Erin Cooley, et al., "Reducing Implicit Racial Preferences: II. Intervention Effectiveness across Time," *Journal of Experimental Psychology: General* 145 (8) (2016): 1001.
- ⁴⁷ Patrick S. Forscher, Calvin K. Lai, Jordan R. Axt, et al., "A Meta-Analysis of Procedures to Change Implicit Measures," *Journal of Personality and Social Psychology* 117 (3) (2019): 522.
- ⁴⁸ Elizabeth Levy Paluck and Donald P. Green, "Prejudice Reduction: What Works? A Review and Assessment of Research and Practice," *Annual Review of Psychology* 60 (2009): 339–367.
- ⁴⁹ Zachary T. Kalinoski, Debra Steele-Johnson, Elizabeth J. Peyton, et al., "A Meta-Analytic Evaluation of Diversity Training Outcomes," *Journal of Organizational Behavior* 34 (8) (2013): 1076–1104.
- ⁵⁰ Katerina Bezrukova, Chester S. Spell, Jamie L. Perry, and Karen A. Jehn, "A Meta-Analytical Integration of Over 40 Years of Research on Diversity Training Evaluation," *Psychological Bulletin* 142 (11) (2016): 1227.
- ⁵¹ Alexandra Kalev, Frank Dobbin, and Erin Kelly, "Best Practices or Best Guesses? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies," *American Sociological Review* 71 (4) (2006): 589–617. See also Alexandra Kalev and Frank Dobbin, "Re-tooling Career Systems to Fight Workplace Bias: Evidence from U.S. Corporations,"

- Dædalus* 153 (1) (Winter 2024): 213–230, <https://www.amacad.org/publication/retooling-career-systems-fight-workplace-bias-evidence-us-corporations>.
- ⁵² Elizabeth Levy Paluck, Roni Porat, Chelsey S. Clark, and Donald P. Green, “Prejudice Reduction: Progress and Challenges,” *Annual Review of Psychology* 72 (2021): 533–560.
- ⁵³ CBS News, “We Asked 155 Police Departments about Their Racial Bias Training. Here’s What They Told Us,” August 7, 2019, <https://www.cbsnews.com/news/racial-bias-training-de-escalation-training-policing-in-america>.
- ⁵⁴ Disclosure: I have served as a paid expert for the lead plaintiffs in the New York City lawsuit.
- ⁵⁵ Robert E. Worden, Sarah J. McLean, Robin S. Engel, et al., *The Impacts of Implicit Bias Awareness Training in the NYPD* (Cincinnati: The John F. Finn Institute for Public Safety, Inc. and The Center for Police Research and Policy at the University of Cincinnati, 2020).
- ⁵⁶ Calvin K. Lai and Jaclyn A. Lisnek, “The Impact of Implicit-Bias-Oriented Diversity Training on Police Officers’ Beliefs, Motivations, and Actions,” *Psychological Science* 34 (4) (2023): 424–434.
- ⁵⁷ Amanda Charbonneau and Jack Glaser, “Suspicion and Discretion in Policing: How Laws and Policies Contribute to Inequity,” *UC Irvine Law Review* 11 (2020): 1327; and Herman Goldstein, “Police Discretion: The Ideal Versus the Real,” *Public Administration Review* 23 (3) (1963): 140–148.
- ⁵⁸ Charles G. Lord, Lee Ross, and Mark R. Lepper, “Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence,” *Journal of Personality and Social Psychology* 37 (11) (1979): 2098.
- ⁵⁹ Harry J. Holzer, Steven Raphael, and Michael A. Stoll, “Perceived Criminality, Criminal Background Checks, and the Racial Hiring Practices of Employers,” *The Journal of Law and Economics* 49 (2) (2006): 451–480.
- ⁶⁰ Abigail Wozniak, “Discrimination and the Effects of Drug Testing on Black Employment,” *Review of Economics and Statistics* 97 (3) (2015): 548–566.
- ⁶¹ Russell J. Skiba, Mariella I. Arredondo, and Natasha T. Williams, “More than a Metaphor: The Contribution of Exclusionary Discipline to a School-to-Prison Pipeline,” *Equity & Excellence in Education* 47 (4) (2014): 546–564; and Erik J. Girvan, Cody Gion, Kent McIntosh, and Keith Smolkowski, “The Relative Contribution of Subjective Office Referrals to Racial Disproportionality in School Discipline,” *School Psychology Quarterly* 32 (3) (2017): 392.
- ⁶² Charbonneau and Glaser, “Suspicion and Discretion in Policing.”
- ⁶³ Ian Ayres, “Outcome Tests of Racial Disparities in Police Practices,” *Justice Research and Policy* 4 (1–2) (2002): 131–142.
- ⁶⁴ Deborah Ramirez, Jennifer Hoopes, and Tara Lai Quinlan, “Defining Racial Profiling in a Post-September 11 World,” *American Criminal Law Review* 40 (2003): 1195; and Malcolm Gladwell, “Troublemakers: What Pitbulls Can Teach Us about Profiling,” *The New Yorker*, January 29, 2006.
- ⁶⁵ Disclosure: I have served as a paid expert for the Floyd plaintiffs during the remedy phase of the NYPD SQF lawsuit.
- ⁶⁶ Disclosure: I have served as a paid expert consultant to the California Department of Justice, advising on the analysis and interpretation of RIPA data.

- ⁶⁷ Charbonneau and Glaser, “Suspicion and Discretion in Policing.”
- ⁶⁸ John MacDonald, Jeffrey Fagan, and Amanda Geller, “The Effects of Local Police Surges on Crime and Arrests in New York City,” *PLOS ONE* 11 (6) (2016): e0157223.
- ⁶⁹ Kyle Smith, “We Were Wrong about Stop and Frisk,” *The National Review*, January 1, 2018, <https://www.nationalreview.com/2018/01/new-york-city-stop-and-frisk-crime-decline-conservatives-wrong/>; and Anthony Fisher, “NY Daily News Admits ‘We Were Wrong’ about Stop and Frisk,” *Reason Magazine*, August 9, 2016.
- ⁷⁰ John M. Darley and Paget H. Gross, “A Hypothesis-Confirming Bias in Labeling Effects,” *Journal of Personality and Social Psychology* 44 (1) (1983): 20.
- ⁷¹ Marianne Bertrand and Sendhil Mullainathan, “Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination,” *American Economic Review* 94 (4) (2004): 991–1013; Holzer, Raphael, and Stoll, “Perceived Criminality, Criminal Background Checks, and the Racial Hiring Practices of Employers”; and Wozniak, “Discrimination and the Effects of Drug Testing on Black Employment.”
- ⁷² Jennifer T. Kubota, “Uncovering Implicit Racial Bias in the Brain: The Past, Present & Future,” *Daedalus* 153 (1) (Winter 2024): 84–105, <https://www.amacad.org/publication/uncovering-implicit-racial-bias-brain-past-present-future/>; and Elizabeth A. Phelps, Kevin J. O’Connor, William A. Cunningham, et al., “Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation,” *Journal of Cognitive Neuroscience* 12 (5) (2000): 729–738.
- ⁷³ Michael Sierra-Arévalo, “American Policing and the Danger Imperative,” *Law & Society Review* 55 (1) (2021): 70–103.
- ⁷⁴ Worden, McLean, Engel, et al., “The Impacts of Implicit Bias Awareness Training in the NYPD”; Lai and Lisnek, “The Impact of Implicit-Bias-Oriented Diversity Training on Police Officers’ Beliefs, Motivations, and Actions”; and California Assembly Bill 392 (2019).
- ⁷⁵ Robert Rogers, “Use of Deadly Force by Police Disappears on Richmond Streets,” *East Bay Times*, September 6, 2014.
- ⁷⁶ Robin S. Engel, Nicholas Corsaro, Gabrielle T. Isaza, and Hannah D. McManus, “Assessing the Impact of De-Escalation Training on Police Behavior: Reducing Police Use of Force in the Louisville, KY Metro Police Department,” *Criminology & Public Policy* 21 (2) (2022): 199–233; and Robin S. Engel, Hannah D. McManus, and Tamara D. Herold, “Does De-Escalation Training Work? A Systematic Review and Call for Evidence in Police Use-of-Force Reform,” *Criminology & Public Policy* 19 (3) (2020): 721–759.
- ⁷⁷ Geller, Goff, Lloyd, et al., “Measuring Racial Disparities in Police Use of Force.”
- ⁷⁸ Hetey, Hamedani, Markus, and Eberhardt, “When the Cruiser Lights Come On.”
- ⁷⁹ Ibid.; Manuel J. Galvan and B. Keith Payne, “Implicit Bias as a Cognitive Manifestation of Systemic Racism,” *Daedalus* 153 (1) (Winter 2024): 106–122, <https://www.amacad.org/publication/disrupting-effects-implicit-bias-case-discretion-policing/>; and John T. Jost and Mahzarin R. Banaji, “The Role of Stereotyping in System-Justification and the Production of False Consciousness,” *British Journal of Social Psychology* 33 (1) (1994): 1–27.