per year in the United States, a thorough accounting by ⸱ ⸱ ⸱ ⸱ ⸱ (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 this 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

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   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] **There**kiuz/doi.hough/coinTO.g 0 115.3   /Prr (plicit bias scores are indirect, representing response speed differences to varied series of stimuli, they correlate with measures that, although subject to self-

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 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.

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 internally 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]

conscious awareness, and occur automatically. That said, implicit bias is not the

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
.2 Wmade cnwisons of searªë ÊìÖ„cEMC   /P <<c/1 (er)2dd8tBT 0 EMC   /P <</L.

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

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 ʆ ، ، .[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 ˷ ، ، ، . 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

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

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.
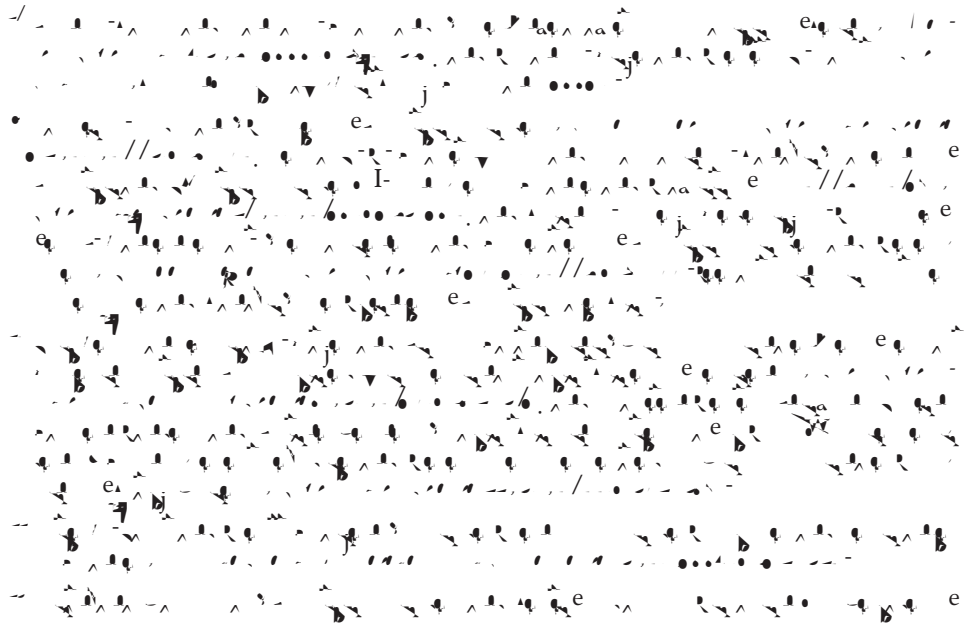

Jack Glaser