

Implicit Bias Training for Police

April 23, 2021

Benjamin Feigenberg

Jack Glaser

University of Illinois at Chicago

University of California, Berkeley

Eleni Packis

University of Chicago Crime Lab

1 Introduction

Research by social psychologists indicates that implicit biases are likely an important driver of discrimination, particularly in the context of policing. Implicit bias trainings are intended to help individuals to confront and reduce their inner, non-conscious biases. In recent years, a significant number of police departments have implemented implicit bias trainings to address concerns regarding how implicit racial biases among officers may contribute to racial disparities in stop, search, use of force, and arrest rates. In 2019, CBS News surveyed 150 large police departments across the country and found that 69% were already implementing implicit bias trainings. However, the majority of departments had not established a methodology for evaluating the efficacy of these trainings. This finding mirrors the broader lack of evidence related to the efficacy of implicit bias training in the context of policing (Smith, 2015).

Below, we first summarize the extensive body of work related to the efficacy of implicit bias interventions in non-policing contexts. We recognize that policing is a unique occupation, and so evidence from other sectors is not a perfect substitute for evidence from the policing sector specifically. Given the limited data on this type of training in the policing context, however, information from other contexts is surely relevant. The general conclusion from this literature is that one-time interventions generate, at most, temporary reductions in implicit biases.¹

We then discuss what is (to the best of our knowledge) the only completed research that links variation in exposure to implicit bias training with real-world policing outcomes: a single study based on a randomized trial that directly examined the link between the introduction of implicit bias training and policing outcomes in the context of New York City. The authors of the study identified modest, short-lived effects of training on the *recognition* of bias. However, the authors found essentially no changes in racial disparities across a range of policing outcomes.

From this available evidence we would *not* conclude that implicit bias training is intrinsically incapable of changing policing outcomes. It is logically possible that other types of implicit bias training content or delivery modalities, or more intensive trainings could have more pronounced impacts than what we have seen to date in both policing and non-policing contexts. But it would appear that the specific implicit bias interventions which have been rigorously

¹ See the Methods for Research Review report for a summary of the criteria used to assess the methodological rigor of existing research and to determine which prior studies to discuss in detail in the present report.

studied are unlikely to have substantial impacts on the policing outcomes of greatest public policy concern.

2 Implicit Bias and its Relation to Behavior

Implicit bias is a concept introduced by social psychologists (e.g., Greenwald and Banaji, 1995) building on decades of cognitive psychological science demonstrating that much of human memory (encoding, storage, and retrieval) operates outside of conscious awareness and can be activated automatically and uncontrollably (e.g., Neely, 1977). Social psychologists have developed computer-based reaction time measures, most prominently the Implicit Association Test (IAT; Greenwald et al., 1998) to estimate implicit biases as a function of the relative facility (indexed by speed) that people have categorizing images or words when they are paired one way versus another (e.g., Black-good/White-bad vs. Black-bad/White-good; see www.projectimplicit.org for demonstrations). Automatically activated implicit biases are more likely to influence spontaneous behaviors, while explicit attitudes (i.e., those measured with questionnaires) are more likely to influence relatively controllable behaviors (Dovidio et al., 2002).

Although implicit bias measures are very indirect -- imputing a mental association based on speed of button pressing in response to randomly displayed stimuli -- they have been found to correlate with explicit measures of bias (Nosek, 2005) as well as with important, organizationally relevant discriminatory behaviors (Jost et al., 2009). Estimated correlations between implicit (and, for that matter, explicit) lab-based bias measures and lab-based discriminatory behaviors are generally modest but have held up in meta-analyses (Oswald et al., 2013, Kurdi et al., 2019). Greenwald et al. (2015) emphasize that small effects on behavior matter in the cumulative sense, across many affected individuals and within individuals' lives over time.

3 Interventions to Address Implicit Bias in Non-policing Contexts

The term "implicit bias training" is used to refer to a range of different types of training content and modalities that look very different in practice. Police implicit bias trainings, as well as those found in other industries, typically involve multiple hours of lecture and discussion on the topic. To inform the design of these real-world trainings and to provide proof-of-concept regarding the malleability of implicit biases, social psychological researchers have attempted to reduce implicit biases using a wide variety of short, strategy-based interventions. While all such interventions have the same underlying aim, to reduce the expression of implicit biases, they employ a variety of different approaches to do so:

- Some aim to reduce biases by promoting *perspective-taking*. These interventions have participants consider the thoughts and behaviors of outgroup members based on the logic that discrimination may result in part from a lack of empathy.
- Other interventions focus on providing exposure to *counter-stereotypic stimuli* based on the theory that increasing counter-stereotypic associations will reduce the manifestation of biases. Counter-stereotypic exposure interventions include providing participants with counter-stereotypical examples in the context of practice IAT exercises and providing counter-stereotypical scenarios in which outgroup members behave well and in-group members behave poorly.

- Another group of interventions focused on *egalitarianism* aim to promote equitable behaviors by, for example, asking participants to recall instances in which they behaved objectively or by providing visual examples of in-group and outgroup members happily engaged in collective behaviors.
- A final broad category of interventions uses *evaluative conditioning*, which aims to reduce biases by promoting the association between outgroup members and positive stimuli. Corresponding interventions include the use of computer-based stimuli to strengthen these associations (for instance, between Black people and “good”).

Two large-scale comparative investigations (Lai et al., 2014, Lai et al., 2016) report carefully controlled tests of a broad range of one-time interventions employing these alternative approaches and specific designs that include those summarized above. In the 2014 study, seventeen interventions, carried out by a coordinated team of research labs in North America and Europe, were tested. More than 17,000 volunteer participants were randomly assigned to trainings (or to a control group) and efficacy was assessed using IATs as well as self-reports. Each intervention was tested in multiple labs, and the results were meta-analyzed to provide robust estimates. Each intervention was designed to take five minutes or less to administer. The authors found that a subset of eight interventions focused on providing counter-stereotypical examples, evaluative conditioning, or very specific tactics (for instance, asking participants to commit to reducing biased responses on the IAT) were effective in reducing implicit bias scores, while those that emphasized perspective-taking and egalitarianism were not. The 2016 follow-up study replicated the bias reductions in the more promising interventions but showed the reductions had completely dissipated when follow-up measures were collected twenty-four hours after the interventions ended. In sum, even the subset of interventions that proved most effective in generating immediate reductions in bias did not generate lasting impacts.

There is some evidence that longer-duration and more intensive interventions (as compared to those tested by social psychological researchers in the meta-analyses described above) may lead to more persistent changes in measured biases. In Devine et al. (2012), the authors implemented a multifaceted, interactive intervention focusing on breaking habits related to implicit biases, spread out over several weeks and lasting over one hour in total (including education, training, and the elicitation of answers to free-responses questions). This small-scale study, which randomly assigned 91 non-Black psychology students to treatment or control groups, found that the intervention went beyond reducing the treatment group’s implicit racial biases and that treatment participants reported greater concerns about racial discrimination and explicit recognition of racial biases that persisted beyond the end of the training. However, the implicit bias reduction results did not replicate in a subsequent study with non-student participants (Carnes et al., 2015).

4 Interventions to Address Implicit Bias in Policing

To measure the effects of implicit bias in a manner more directly applicable to the policing context, psychologists have developed a “shooter task.” This task estimates unintended racial bias by comparing response latencies (the time it takes someone to make the decision about whether to shoot or not) and error rates (whether the person makes the right decision to shoot or not, depending on whether the other subject in the exercise is armed or not) in a computer-based simulation in which participants make a speeded decision to shoot or not shoot

armed/unarmed male targets (Correll et al., 2002, Mekawi and Bresin, 2015).² Racial bias is assessed by randomly varying the race of the male targets in these simulations, with the usual study contrast focusing on comparisons of Black and White male targets.

Shooter bias has been shown to be correlated with an implicit race-weapon stereotype (whereby Blacks are more readily associated with weapons and Whites are more readily associated with tools), but not with the more global race-prejudice implicit bias (whereby Blacks are more readily associated with “bad” stimuli and Whites are more readily associated with “good” stimuli) (Glaser and Knowles, 2008). One early “shooter task”-based study with a sample of 48 police officers found that shooter bias (the tendency to make the errors of shooting unarmed Blacks more than unarmed Whites and failing to shoot armed Whites more than armed Blacks) attenuated significantly over time as officers carried out the task when race and being armed were uncorrelated in the simulation. However, the study could not examine whether the intervention affected officers’ decision making in the field (Plant and Peruche, 2005).

Another study in this space demonstrated that two large samples of police officers displayed similar response time shooter biases (i.e., faster shooting of armed Blacks than armed Whites, and faster non-shooting of unarmed Whites than unarmed Blacks) to community (i.e., general public) samples, but in contrast to the general public police did not show bias in terms of making errors – shooting unarmed targets (Correll et al., 2007). Relatedly, James et al. (2016) evaluated officer behavior in video-based simulations in which officers were given considerably more time to make shoot/don’t-shoot decisions. In that setting, the authors found that officers were actually slower to shoot armed Blacks than armed Whites and less likely to shoot unarmed Black suspects than unarmed White suspects, despite exhibiting overwhelming rates of Black-weapon association in an Implicit Association Test. Similarly, Park et al. (2008) found that research participants (college students) exhibited higher levels of shooter bias when their cognitive resources were depleted, suggesting that, even in the speeded task, some control can be effectively applied. These findings highlight the importance of automaticity in driving the shooter bias identified in prior work; given the lengthiness of the simulation scenarios, officers in the James et al. (2016) study had time to engage in controlled processes and were likely able to marshal the motivation and behavioral intention to avoid erroneously shooting unarmed Blacks. Importantly, to the best of our knowledge, these “shooter task” measures have not been validated as predictors of real-world policing use of force decisions, which are, of course, much more complex and psychologically fraught.

While officer-involved shootings are far too common from a civil rights and public policy perspective, from a statistical perspective even in large departments there are often not enough shootings each year to facilitate the sort of statistical analysis that would allow us to test for links with implicit bias-influenced behaviors like shooter bias. More commonplace behaviors like vehicle and pedestrian stops and searches and use of nonlethal force can be analyzed for racial disparities with considerably greater statistical power. Nevertheless, little rigorous

² The nature of stereotypic associations between Blacks and crime among police officers has also been explored using a variety of alternative study designs (see, for instance, Eberhardt et al., 2004).

research has been conducted on the effectiveness of implicit bias trainings for police officers with respect to field performance.³

More fundamentally, there is reason to be skeptical that existing classroom-based trainings will reduce the impact of implicit bias on real world decision making. First, it is not clear to what degree such trainings can influence real-world behaviors (as opposed to short-term lab-based measures) given the fundamental nature of implicit bias—which is that it operates outside of conscious awareness and control, reflecting strong mental associations that are learned over long periods of time, reinforced by extant environments, and influence judgments without subjective experience (Spencer et al., 2016). Second, on a practical level, the lecture-based structure of existing police department trainings may further reduce the likelihood that these trainings can influence real-world policing behaviors given the lack of emphasis placed on repeated practice.

Recent survey-based evidence from evaluations of implicit bias trainings does provide some cause for optimism. For instance, the Urban Institute’s evaluation of interventions developed through the National Initiative for Building Community Trust and Justice found that, after implicit bias trainings, officers tended to have greater understanding of the potential for implicit biases and situational factors to influence their decisions. However, the evaluation did not examine whether these understandings led to changes in officers’ behaviors in the field. Moreover, since these trainings were conducted in tandem with the introduction of other new policies and trainings, it is difficult to isolate the impacts of implicit bias trainings from the complementary trainings in procedural justice and efforts to promote community policing (Jannetta et al., 2019).⁴

To date, the only available evidence of which we are aware on the real-world policing impacts of implicit bias training comes from a recently completed large scale study, conducted with the New York City Police Department. This study represents a first attempt to rigorously link implicit bias training to policing outcomes. As part of this research, approximately 15,000 officers in the Patrol Services Bureau, Transit Bureau and Housing Bureau received the most widely employed training in implicit bias at randomized, pre-specified points in time between May 2018 and April 2019. This training, based on the Fair and Impartial Policing (FIP) curriculum, emphasizes bias awareness, highlights the potential consequences of biased policing, and provides tools that can be used to manage bias in the field. Specifically, officers are taught that biases are widespread even among well-intentioned individuals and that implicit biases can impact behaviors without conscious awareness in ways that make policing “unsafe, ineffective, and unjust.” The strategies that officers are encouraged to adopt to manage biases include slowing down their responses to reduce ambiguity and conducting “self-checks” on behaviors that may be influenced by biases. The training includes lectures as well as interactive activities and lasts for eight hours in total (Worden et al., 2020). As such, the training duration is markedly

³ More generally, a 2017 report by the National Academies of Sciences, Engineering, and Medicine states that “There is, to our knowledge, no peer-reviewed work in psychology examining how any motivating factors, implicit or explicit, held by police influence their behavior toward subjects in the real world” (NASEM, 2017).

⁴ Similar survey-based findings are reported in other settings where officers received bundled trainings that included implicit bias components. See, for instance, Stanford SPARQ (2015), which reports pre-post survey results for California law enforcement leaders who participated in a one-day training in 2015.

longer than the durations of the interventions that have been rigorously tested by social psychological researchers and are described above.

To analyze the impacts of the implicit bias training, a team of researchers utilized a number of complementary data sources. First, the authors collected survey data related to “beliefs and attitudes relating to stereotypes and prejudice” either before or immediately after the conclusion of the training (with survey timing randomized across officers so that each officer responded only before or only after the training). Across survey-based indices, the authors consistently found that post-training responses were different from pre-training responses. However, the largest measured difference was only a 0.35-point change (on a five-point scale) for the index composed of eight questions related to beliefs about implicit bias.⁵ In sum, the authors identified modest effects of training on the recognition of bias. These effects appeared to decay over time based on a subsequent follow up survey. It is worth noting that these survey-based estimates are subject to standard social desirability bias concerns (the concern that respondents may simply be telling the interviewer what they think the interviewer wants to hear – the “socially desirable” answer). In addition, there are two concerns related to survey response rates. First, those surveyed before training had a 65.7% response rate, while those surveyed immediately after training had an 85.6% response rate. This difference in response rates may lead measured differences in outcomes as a function of survey timing to confound training treatment effects with compositional differences across the two groups in who chose to respond to the surveys. Second, the low (10%) response rate for the authors’ follow-up survey raises concerns regarding the representativeness of respondents. It is important to note that neither of the surveys exploited the randomized rollout of the training treatment across police districts. The training day surveys were essentially based on a pre- versus post-training comparison (with the researchers conducting only a pre- *or* post-training survey with any given officer). The follow-up survey was conducted after all participating officers had already been trained.

To estimate the effects of the training on policing outcomes, the authors did exploit the randomization of training timing across 10 district clusters (with each district cluster assigned to begin training at a distinct date/step). Based on this design, at each point in time, the authors could compare outcomes in the subset of clusters where training had already taken place to outcomes in the subset of clusters where training had not yet taken place. Since training timing was randomized across clusters, differences in outcomes between trained and untrained clusters at each point in time could thus be causally attributed to the training. Overall, 93.1% of officers were trained with the cluster to which they were assigned, and the authors provide evidence that officer reassignment subsequent to training was limited. However, there is some evidence of imbalance by treatment assignment with respect to the race and ethnicity of those arrested, stopped, and receiving summonses in the pre-intervention period. This “imbalance” by treatment assignment is a concern because it suggests that random assignment may not have fulfilled its primary goal of constructing “treatment” and “control” groups that are similar on average in all respects other than receipt of the intervention itself, which is what enables experiments normally to be able to isolate the effect of the intervention from other confounding factors. It is also worth noting that the rollout of implicit bias training took place during a period in which multiple

⁵ One question, for example, asks whether respondents agree or disagree with the statement that “Implicit biases can affect even individuals who consciously reject prejudices and stereotypes.” Responses are coded on a five-point scale from 1 (“Strongly disagree”) to 5 (“Strongly agree”).

reforms were being implemented within NYPD, including the deployment of body-worn cameras and the relaxation of existing marijuana enforcement policies.

To analyze the impact of training on policing performance, the authors utilized event-level data from NYPD on arrests, stops, frisks in stops, summonses, searches in stops, use of force in stops, use of force in arrests, and citizen complaints. They employed a benchmark specification that was a Poisson regression of the police cluster-by-step level count corresponding to a given event outcome on an indicator for the implicit bias training treatment. Each regression also included police cluster and step fixed effects and the authors constructed heteroskedasticity-robust standard errors. The authors further disaggregated the data and employed the same specification to examine how treatment effects varied based on citizen race and ethnicity.

The authors found essentially no changes in racial disparities across the range of policing outcomes listed above (Worden et al., 2020). Estimated treatment effects were of inconsistent sign (some implying, if taken literally, beneficial effects, others taken at face value implying adverse effects). For instance, treatment effects for overall outcome changes measured across all sample events (and all citizens) ranged from a 5.4% decline in searches to a 16.7% increase in use of force in stops; the only marginally significant overall treatment effect was a 7.2% increase in summonses. More generally, the small number of significant estimates reported were also inconsistent in sign and may be explained by the large number of hypotheses tested. As summarized in Figure 1, which reports estimated percent changes in outcome counts associated with treatment, confidence intervals vary notably across outcomes and are quite large in some cases; on average, 95% confidence intervals can be used to rule out treatment effects greater than roughly 15%.⁶ The authors summarize their findings on enforcement behavior by noting that “The analyses do not lead us to reject the null hypothesis of no training effects on enforcement behavior.”⁷

5 Conclusions

In sum, the only evidence available on the link between implicit bias training and real-world policing outcomes comes from a single study of the randomized rollout of implicit bias training across clusters of NYPD officers, and this study found no evidence that the evaluated training impacted the policing outcomes of greatest public policy concern. As such, further evidence is needed to assess whether trainings that differ in terms of both content and dosage may hold more promise. The evidence described above from social psychological research suggests that longer-duration and more intensive interventions, including those that focus on

⁶ The authors also present supplementary estimates based on a second model that includes an indicator for the September 2018 relaxation of existing marijuana enforcement policies and a third model that interacts this indicator variable with treatment assignment (in practice, the second model appears essentially identical to the first given the inclusion of step fixed effects). The authors also include logit specifications that examine treatment-induced changes in the likelihood that a given event involves a particular officer behavior and that directly examine whether estimated treatment effects vary by whether the subject of a given event is Black or Hispanic. The authors also examine treatment effect heterogeneity by officers’ training day survey responses.

⁷ It should be noted that the raw number of stops and frisks during the period studied was a small fraction of the number at the peak of Stop & Frisk in New York, leaving less opportunity for improvements in racial disparities.

repeated practice, may prove more effective. This hypothesis is reinforced by meta-analyses from related fields.⁸

Some also hypothesize that combating implicit biases at the supervisory level (i.e., among sergeants and other higher-ranked officers) may prove equally or more effective in improving the outcomes of police-citizen interactions. Although rigorous research on the role of sergeants in influencing police-citizen interactions is scarce, there is a widespread belief that sergeants play a critical role in establishing departmental norms (PERF, 2018). Moreover, sergeants and other higher-ranked supervisors establish officer incentives. To the extent that these incentives motivate officers to increase stop counts, for instance, the marginal stops conducted by officers may leave them more vulnerable to the influence of bias. Understanding how officer incentives interact with implicit biases represents a worthwhile direction for future research.

References

- Bezrukova, K., C. Spell, J. Perry, and K. Jehn. 2016. "A Meta-analytical Integration of Over 40 Years of Research on Diversity Training Evaluation." *Psychological Bulletin* 142(11), 1227–1274.
- Carnes, Molly, Patricia G. Devine, Linda Baier Manwell, Angela Byars-Winston, Eve Fine, Cecilia E. Ford, Patrick Forscher, Carol Isaac, Anna Kaatz, Wairimu Magua, Mari Palta, and Jennifer Sheridan. 2015. "Effect of an Intervention to Break the Gender Bias Habit for Faculty at one Institution: A Cluster Randomized, Controlled Trial." *Academic Medicine: Journal of the Association of American Medical Colleges* 90(2), 221-230.
- Correll, Joshua, Bernadette Park, Charles M. Judd, and Bernd Wittenbrink. 2002. "The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals." *Journal of Personality and Social Psychology* 83(6), 1314-1329.
- Correll, Joshua, Bernadette Park, Charles M. Judd, Bernd Wittenbrink, Melody Sadler, and Tracie Keese. 2007. "Across the Thin Blue Line: Police Officers and Racial Bias in the Decision to Shoot." *Journal of Personality and Social Psychology* 92(6), 1006-1023.
- Devine, Patricia G., Patrick S. Forscher, Anthony J. Austin, and William T.L. Cox. 2012. "Long-term Reduction in Implicit Race Bias: A Prejudice Habit-breaking Intervention." *Journal of Experimental Social Psychology* 48(6), 1267-1278.
- Dovidio, John F., Kerry Kawakami, and Samuel L. Gaertner. 2002. "Implicit and Explicit Prejudice and Interracial Interaction." *Journal of Personality and Social Psychology* 82(1), 62-68.

⁸ See, for instance, Bezrukova et al. (2016), which conducts a meta-analysis of diversity training impacts and concludes that longer-duration trainings are more effective.

Eberhardt, Jennifer L., Phillip A. Goff, Valerie J. Purdie, and Paul G. Davies. 2004. "Seeing Black: Race, Crime, and Visual Processing." *Journal of Personality and Social Psychology* 87(6), 876-893.

Glaser, J., and E. D. Knowles. 2008. "Implicit Motivation to Control Prejudice." *Journal of Experimental Social Psychology* 44(1), 164-172.

Greenwald, Anthony G., and Mahzarin R. Banaji. 1995. "Implicit Social Cognition: Attitudes, Self-esteem, and Stereotypes." *Psychological Review* 102(1): 4-27.

Greenwald, Anthony G., Mahzarin R. Banaji, and B.A. Nosek. 2015. "Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects." *Journal of Personality and Social Psychology* 108, 553-561.

Greenwald, Anthony G., Debbie E. McGhee, and Jordan L.K. Schwartz. 1998. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology* 74(6), 1464-1480.

James, Lois, Stephen M. James, and Bryan J. Villa. 2016. "The Reverse Racism Effect: Are Cops More Hesitant to Shoot Black Than White Suspects?" *Criminology and Public Policy* 15(2), 457-479.

Jannetta, Jesse, Sino Esthappan, Jocelyn Fontaine, Mathew Lynch, and Nancy La Vigne. 2019. "Learning to Build Police-Community Trust." The Urban Institute.

Jost, J.T., L. Rudman, I.V. Blair, D.R. Carney, N. Dasgupta, J. Glaser, and C. Hardin. 2009. "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, 39-69.

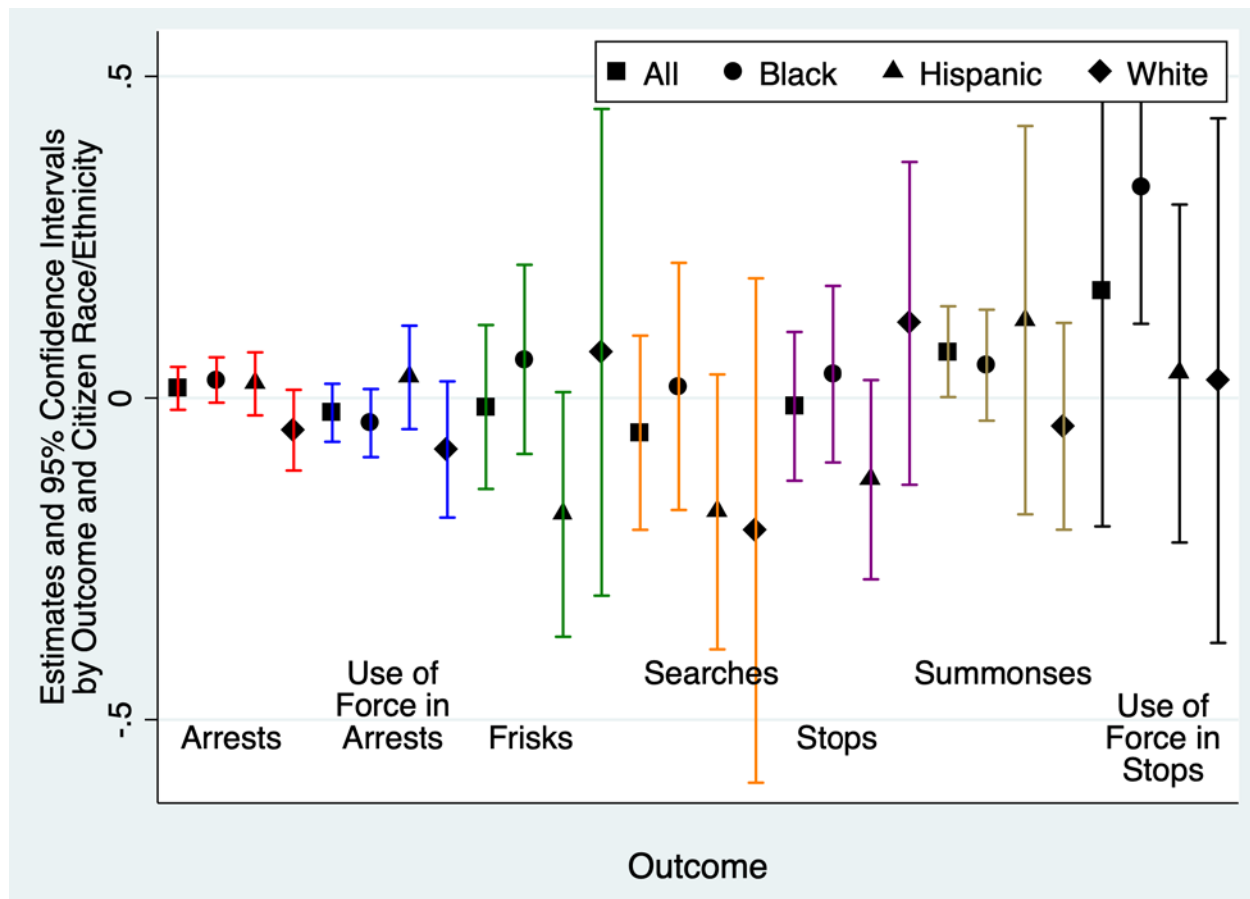
Kurdi, Benedek, Allison E. Seitchik, Jordan R. Axt, Timothy J. Carroll, Arpi Karapetyan, Neela Kaushik, Diana Tomezsko, Anthony G. Greenwald, and Mahzarin R. Banaji. 2019. "Relationship Between the Implicit Association Test and Intergroup Behavior: A Meta-analysis." *American Psychologist* 74(5), 569-586.

Lai, Calvin K., Maddalena Marini, Steven A. Lehr, Carlo Cerruti, Jiyun-Elizabeth L. Shin, Jennifer A. Joy-Gaba, Arnold K. Ho, et al. 2014. "Reducing Implicit Racial Preferences: I. A Comparative Investigation of 17 Interventions." *Journal of Experimental Psychology: General* 143(4), 1765-1785.

Lai, Calvin K., Allison L. Skinner, Erin Cooley, Sohad Murrar, Markus Brauer, Thierry Devos, Jimmy Calanchini, et al. 2016. "Reducing Implicit Racial Preferences: II. Intervention Effectiveness Across Time." *Journal of Experimental Psychology: General* 145(8), 1001-1016.

- Mekawi, Yara, and Konrad Bresin. 2015. "Is the Evidence from Racial Bias Shooting Task Studies a Smoking Gun? Results from a Meta-analysis." *Journal of Experimental Social Psychology* 61, 120-130.
- National Academies of Sciences, Engineering, and Medicine (NASEM). 2017. "Proactive Policing: Effects on Crime and Communities." Washington, DC: The National Academies Press.
- Neely, James H. 1977. "Semantic Priming and Retrieval from Lexical Memory: Roles of Inhibitionless Spreading Activation and Limited-capacity Attention." *Journal of Experimental Psychology: General* 106(3), 226-254.
- Nosek, Brian A. 2005. "Moderators of the Relationship Between Implicit and Explicit Evaluation." *Journal of Experimental Psychology: General* 134(4), 565-584.
- Oswald, Frederick L., Gregory Mitchell, Hart Blanton, James Jaccard, and Philip E. Tetlock. 2013. "Predicting Ethnic and Racial Discrimination: A Meta-analysis of IAT Criterion Studies." *Journal of Personality and Social Psychology* 105(2), 171-192.
- Park, Sang Hee, Jack Glaser, and Eric D. Knowles. 2008. "Implicit Motivation to Control Prejudice Moderates the Effect of Cognitive Depletion on Unintended Discrimination." *Social Cognition* 26(4), 401-419.
- Plant, E. Ashby, and B. Michelle Peruche. 2005. "The Consequences of Race for Police Officers' Responses to Criminal Suspects." *Psychological Science* 16(3), 180-183.
- Police Executive Research Forum (PERF). 2018. "Promoting Excellence in First-Line Supervision: New Approaches to Selection, Training, and Leadership Development."
- Smith, Robert J. 2015. "Reducing Racially Disparate Policing Outcomes: Is Implicit Bias Training the Answer?" *U. Haw. L. Rev.* 37, 295.
- Spencer, Katherine B., Amanda K. Charbonneau, and Jack Glaser. 2016. "Implicit Bias and Policing." *Social and Personality Psychology Compass* 10(1), 50-63.
- Stanford SPARQ. 2015. "Principled Policing: Procedural Justice and Implicit Bias Training." White paper.
- Worden, Robert E., Sarah J. McLean, Robin S. Engel, Hannah Cochran, Nicholas Corsaro, Danielle Reynolds, Cynthia J. Najdowski, and Gabrielle T. Isaza. 2020. "The Impacts of Implicit Bias Awareness Training in the NYPD." The John F. Finn Institute for Public Safety, Inc.

Figure 1: Worden et al. (2020) Estimates of Percent Changes in Policing Outcomes by Outcome and Citizen Race/Ethnicity



This figure displays point estimates and 95% confidence intervals from Poisson regressions of police cluster-by-step level counts corresponding to the specified outcomes on an indicator for the implicit bias training treatment. Each regression also includes police cluster and step fixed effects and presents heteroskedasticity-robust standard errors. Race/ethnicity-specific estimates reflect percent changes in counts of the relevant outcome for members of that group (i.e., the percent change in arrests of Blacks at the police cluster-by-step level).