Anti-bias training and perceived force climate: links with prejudiced attitudes in United Kingdom policing

Article

Accepted Version

In press: *Journal of Community & Applied Social Psychology*

Anti-bias training and perceived force climate: Links with prejudiced attitudes in United Kingdom policing

Illinois Institute of Technology

Netta Weinstein*

University of Reading

Les Graham

Durham University

Marisa Plater

Durham University

Author note

*co-first authors
Corresponding author: Nicole Legate, Department of Psychology, Illinois Institute of Technology, 3424 S. State St., Chicago IL, 60616, 312.567.3500, nlegate@iit.edu.
Conflict of interest statement: The authors have no conflicts of interest to report.
Funding statement: This work was not supported by external funding.
Data availability and open science statement: Preregistration and materials can be found on our study page: https://osf.io/hbja3/?view_only=8e34105ad15c47299c4f30a12430d75d. Data cannot be made available due to restrictions within UK policing.

Abstract

Anti-bias training has been viewed as the solution to prejudice in organizations, yet the evidence is mixed in real-world settings. Some point to the broader organizational climate that training takes place in as critical, and herein we investigate one aspect: communicating about bias in autonomy-supportive (i.e., non-shaming) ways. Using the 2019 National Well-Being and Inclusion Survey of UK police officers and staff (n=34,529 in 43 forces), we tested links of participating in anti-bias training, perceived autonomy-supportive communication, and their interaction on prejudiced attitudes. Results revealed a negligible effect (R^2=0.001) of participating in anti-bias training but a moderate effect (R^2=0.05) of perceiving autonomy-supportive communication predicting lower prejudice. Their interaction was significant but negligible (R^2=0.001): participating in anti-bias training predicted lower prejudice when perceiving autonomy-supportive communication; there was no link between training and attitudes without autonomy-supportive communication. Implications for improving the effectiveness of anti-bias training in applied settings and research are discussed.

Keywords: anti-bias training, unconscious bias, motivation, organizational context

Anti-bias training and perceived force climate predicting prejudiced attitudes in UK policing

Recent high-profile incidents of police bias have cast a spotlight on the need to confront the issue of prejudice within policing (Cooper & Fullilove, 2020), but this problem is not new (Chan, 1997), nor is it restricted to the institution of policing (e.g., Oberai & Anand, 2018). To mitigate the problem in policing and other organizations, anti-bias training – a term often used synonymously with diversity training – is offered or mandated for employees. Generally, anti-bias training is designed to reduce biased attitudes and behavior, and often involves psychoeducation, or educational information about the psychological aspects of bias, to help raise awareness of the biases that individuals may hold (HMICFRS, 2017). Anti-bias training as the proposed solution to the problem of prejudice is not specific to policing (Carter et al., 2020), with billions of dollars spent annually on these trainings (Kirkland & Bohnet, 2017). Despite high investment, training to curb prejudice tends to show modest effects immediately following the intervention but few lasting effects (Paluck et al., 2021), failing to translate into real-world gains in inclusivity in workplaces (Chang et al., 2019; Kalev et al., 2006).

The most ubiquitous type of anti-bias training is unconscious bias training (also commonly called implicit bias training), a general approach of making people conscious of their own negative biases, attitudes, and stereotypes, with the idea that people made aware of their own attitudes and behaviors will then adjust them (Atewologun et al., 2018). The evidence for unconscious bias training is mixed and hotly contested (Noon, 2018). Unconscious bias shows weak links with biased behavior (Oswald et al., 2013), and large-scale interventions designed to reduce anti-bias show that at best, any benefits tend to dissipate after a day (Lai et al., 2016). Much of the foundational work evidencing effects was conducted with student samples (Forscher

et al., 2019) with limited evidence in applied settings. Broader reviews of anti-bias and diversity training beyond unconscious bias training finds more positive results, though effects on attitudes are small and tend to decay over time (Bezrukova et al., 2016).

Not all anti-bias training looks the same, and practitioners may be tempted to conclude the context or content of their training defies the odds set out in lab work. In support of this view, a recent review comparing anti-bias training approaches showed that some approaches had greater promise than others (FitzGerald et al., 2019). Defenders further argue that little empirical attention has been given to understanding anti-bias training in real-world settings (Hannah & Carpenter-Song, 2003), where training exists within a larger workplace climate that, ideally, gives it purpose and context (Tannenbaum & Yuki, 1992). Only one study to date has examined the effectiveness of anti-bias training within policing, finding small immediate effects on attitudes towards discrimination (i.e., considering discrimination a social problem) and few effects at follow-up (Worden et al., 2020). This study provided an impressive test of the effectiveness of training in the largest police force in the US, yet questions remain about generalizability, both in terms of the effect of anti-bias training across different forces and different types of training and understanding the broader organizational context that may play a role in prejudice-reduction efforts. Indeed, Worden and colleagues (2020) concluded that police antibias training was likely to have a greater impact when supported by other organizational forces and speculated immediate supervisors may be particularly influential.

Increasingly in this literature, there is recognition that changing the organizational climate is critical for prejudice reduction (Carter et al., 2020; Souhami, 2014). Others further argue that formal interventions only produce attitudinal change when there is a broader workplace culture that supports diversity and inclusion (Applebaum, 2019). In support of this, a

meta-analysis found that outcomes of diversity training improved when organizations integrated training with other diversity initiatives (Bezrukova et al., 2016).

One aspect of the organizational climate that shapes employee motivation stands out as a candidate to potentially impact the effectiveness of anti-bias training. An approach based in self-determination theory (SDT; Deci & Ryan, 2017) argues that individuals respond best when they perceive *autonomy support* from the organization around the issue of bias. Autonomy support involves minimizing pressure and shame and helping people to internalize or come to endorse the importance of change in ways that align with their own values and beliefs. A large literature in SDT, especially within the domain of health behavior change, shows that supporting autonomy is essential to fostering lasting behavior change (Sheeran et al., 2021). Closely related to the present study, a series of lab studies evidence a role for autonomy support in communicating about bias and inclusion in ways that promote internalized motivation and reduce prejudiced attitudes (Legault et al., 2011). Importantly, these experiments showed that communicating about bias in ways that thwart autonomy through shaming and blaming language backfired and actually *increased prejudice*. Communicating about bias using autonomy support motivates people to find the value in, and place importance on, prejudice reduction by providing good reasons to change, and concrete strategies on how to achieve change, while avoiding shaming and pressuring people. Fieldwork in policing shows that forces that provide autonomy support for reducing bias predict less prejudiced attitudes among officers and staff (Authors, 2022). However, this motivational factor has never been examined alongside training; perceiving autonomy support from the force may strengthen links between training and attitudes by motivating change in accord with insights gained in training.

Current Study

We were interested in investigating the link between anti-bias training and attitudes in the institution of policing, where prejudice reduction is of critical importance. Preliminary data we collected within three UK forces (n = 2,538) showed no relationship of participating in unconscious bias or other anti-bias training on antagonism towards investing in diversity, a subtle explicit measure of prejudiced attitudes. Further, we observed a marginal moderation effect such that an effect of training was observed (predicting lower prejudice) when perceiving the force to communicate about bias in autonomy-supportive ways; there was no training-attitude link in the absence of autonomy support (see the study preregistration for more details).

The 2019 National Well-Being and Inclusion Survey of police officers and staff across the United Kingdom (U.K.) provided a unique opportunity to test the link between participating in anti-bias training and prejudiced attitudes with a large sample representing forces across the country. Further, each force in the U.K. is responsible for developing its own training, so data represent 43 distinct implementations within 43 force-specific climates. This sample allows us the power to detect potential moderation effects of employees' perceptions of workplace climate factors that may make individuals more or less responsive to training. These series of tests– sensitive to perceptions of the force and attitudinal correlates of training- allow us to test nuances of the data, an important step in translating research findings into practice and policy change (Ijzerman, et al., 2020).

Given the mixed findings evidenced in the literature and our own preliminary investigation, rather than identifying a directional hypothesis, we sought to understand the overall effect and variability in this effect across police forces within the U.K. Thus, we set out to investigate 1) the strength, direction, and variability across training approaches of the relationship between anti-bias training and prejudiced attitudes and 2) the relationship between

perceived broader workplace support for diversity and inclusion. Specifically, we tested whether prejudiced attitudes would be lower when people have had anti-bias training and they perceive their force to communicate about bias in autonomy-supportive versus controlling ways.

Methods

Participants and Procedure

We detailed our research questions, methods, and analytic strategy on our study preregistration. There are two deviations from our preregistration. First, we used the term "unconscious bias training" in the preregistered hypotheses but changed it to "anti-bias training" in this report as it more accurately reflects what was assessed. Second, we used clarity of consequences for biased actions (e.g., suspensions) as a covariate. Employees may be motivated to inhibit expressing prejudice for fear of the consequences to their job, and how clear this is to employees may be linked with autonomy support as well as prejudiced attitudes, rendering it a useful covariate.

Data came from the 2019 National Well-being and Inclusion Survey of police officers and staff across the U.K. (N = 34,529) after the survey was approved by a university ethics committee (reference: DUBS-2019-10-30T10:50:41-qkvc59). Police forces within the U.K participated in late 2019 and early 2020 and answered questions about their well-being more generally and at work, as well as questions about inclusion and bias at work. For the confirmatory portion of this study, we examine the question asking them whether they took part in unconscious bias or anti-bias training, and attitudes they hold towards various minority groups. Data were collected from officers and staff through force-wide distributions. All participants provided informed consent. To ensure anonymity and due to restrictions placed on the data collected for this study, data cannot be made publicly available.

We calculated the minimum sample size to identify an overall small effect size of $d = .20$ at 95% power ($n = 1{,}084$). Sensitivity analyses indicated that our actual sample size enabled us to detect effect sizes much smaller, $d = .04$. Forty-three territorial police forces in England and Wales (out of 43) participated with samples ranging in size substantially from 88 to 4,557 officers and staff, $M = 796.93$, $SD = 898.22$. Forces represented were anonymized in order to protect the identities of those forces. There was little missing data on the two key variables required for analyses: taking part in anti-bias training and antagonism (across the sample $n = 1{,}169$ or 3.3% did not respond to these items).

Measures

Anti-Bias Training (predictor). Training was measured with the single item: "Have you taken part in an 'unconscious bias' training course or other course meant to understand bias or prejudice within your force?", which was answered with a "Yes" or "No" response. The aggregated sample was roughly split between participants who reported taking an unconscious bias or other anti-bias training (53%, $n = 18{,}183$) and those who did not (47%, $n = 15{,}919$). Forces ranged widely in terms of percentages of respondents who completed training (19% to 85%) but on average, roughly half (49%) of employees with a force participated in training.

Diversity Antagonism (Outcome). Attitudes about diversity were measured with a short scale, Antagonism towards Investing in Diversity (Al-Khouja et al., 2020). Participants indicated agreement (1 = *strongly disagree*, 7 = *strongly agree*) with three statements "about people from minority groups (for example, those of a different race, ethnicity, or religion)". Items were "The force puts too much emphasis on issues concerning minority groups," "Minority groups demand too much from the force" and "Over the past few years people from minority groups have received more attention than they deserve." Items showed good reliability ($\alpha = .90$), and this

scale, including this three-item version, have been used in prior work in policing and show high correlations with other commonly used measures of prejudice, such as the feeling thermometer (Weinstein et al., 2022).

Autonomy Support to Reduce Bias (Moderator). Participants were asked a brief 5-item version of the Autonomy Support to Reduce Bias scale (Weinstein et al., 2022). This scale asks participants how they experience messages within their force about "attitudes and behaviours towards individuals from minority groups, or those who are different from you (for example, those of a different race, ethnicity, or religion)." They indicated their agreement (1 = *strongly disagree* to 7 = *strongly agree*) on five items assessing how much they perceive their force provides autonomy support around the issue of bias reduction (e.g., "The force has clearly communicated the reasons and need for treating individuals from minority groups in a non-biased manner"; "People at work try to make me feel ashamed in order to get me to act without bias towards individuals from minority groups," reverse-coded). Items showed good reliability ($\alpha = .79$), and this scale, including this shortened five-item version, has been used in prior work in policing and shows links with prejudice cross-sectionally and over time (Weinstein et al., 2022).

Clarity of Consequences (Covariate). Enforcing (punitive) consequences is an appropriate way to respond to prejudice in the workplace, including within policing, and one with multiple benefits to the organization and the communities with which it interacts (Lawrence, 2009). Given it is another motivational tactic, we included this as a covariate in analyses. One item assessed the clarity of consequences for discrimination using the same prompt as above (how participants experience messages within their force about "attitudes and behaviours towards individuals from minority groups"). The item, "the consequences of

discriminating against someone from a minority group are clear" was paired with the scale (1 = *strongly disagree* to 7 = *strongly agree*).

Demographics (Covariates). Participants were asked a number of demographic questions related to their personal and professional identities. Sex was dichotomized into female (*n* = 16,357) and male (*n* = 16,735), with those indicating intersex (*n* = 25) and preferring not to say (*n* = 944) excluded[1]. Five age ranges were provided from *18-25 years* (*n* = 2,010) to *56 years and above* (*n* = 3,681). Minority ethnicity categories were *Asian or Asian British* (*n* = 4532), *Black or Black British* (*n* = 133), *Mixed* (*n* = 394), and these were aggregated into a dummy coded BAME group (coded 1) compared with those who were *White (British/Irish/European)*, *n* = 18,849 (coded 2). Those who preferred not to say were excluded from analyses (*n* = 334). Role defined police officers, coded 1 (*n* = 18,066), compared with those in police staff and police community support officer roles, coded 2 (*n* = 16,072). Tenure in policing ranged from 1- *less than a year* (*n* = 1,730) to 5- *over 20 years* (*n* = 9,306), with most indicating a *11-20 year* tenure (*n* = 14,240).

Results

Analytic Approach

Linear mixed models run in the statistical program Jamovi (version 1.6.3) were used with officers and staff (Level-1) nested within 43 forces (Level-2). In a first step, we examined the intraclass correlation coefficients (ICC) of the three outcomes to estimate the variability of diversity antagonism at the individual level and the force level. The ICC for diversity antagonism

[1] Because we were interested in controlling for how antibias training may interact with or differentially relate to those with various demographic characteristics, and because of the relatively small sample sizes of some of the categories (e.g., intersex, Black or Black British), we dichotomized the two variables sex and race/ethnicity. Anyone who did not respond to those two questions, and anyone reporting sex other than male or female, were excluded from the dichotomous variables of sex and race/ethnicity used in analyses. All other demographic variables were treated as continuous in analyses.

was quite low, .017, meaning that only 1.7% of the variance in diversity antagonism was operating at the force level, with almost all of the variance explained at the individual level. Despite this, we proceeded with the linear mixed model as LRT tests showed that these small force level random effects were significant (p = .01). Further, even if effects are small, it is still recommended to account for force-level clustering, especially as we were interested in force-level effects alongside individual-level ones, and even a small ICC can bias standard errors in an ordinary-least squares regression framework (Bliese et al., 2018).

Next, we ran an unadjusted model with anti-bias training as the sole predictor (Level-1) and prejudiced attitudes operationalized in terms of diversity antagonism served as the outcome. We also ran this unadjusted model with anti-bias training as a random effect, allowing us to test whether forces vary in the effect of anti-bias training on attitudes.

Following this, we ran an adjusted model that included moderation effects and covariates (clear consequences, tenure in policing, sex, race, role, and age). This allowed us to test whether the relationship between anti-bias training and attitudes was moderated by perceived autonomy support to reduce bias in the force. All moderation effects and covariates were entered at Level-1, the individual level.

All models used a REML estimator. The dichotomous predictor anti-bias training and demographic items were left uncentered, and all other variables were grand-mean centered. We present findings from both unadjusted and adjusted models. We interpret statistical significance for effects in terms of the confidence interval not crossing 0. Effect sizes were calculated to get R^2 values (Edwards et al., 2008) and are interpreted according to Cohen's (1988) benchmarks. Table 1 shows the effects of unadjusted and adjusted models on diversity antagonism.

Linear Mixed Model Results

To evaluate our research questions, the first linear mixed model we examined was the unadjusted model predicting diversity antagonism from participation in anti-bias training. Results showed that participating in anti-bias training related to lower diversity antagonism, b = -.14 (SE = .02), 95% confidence interval [-.17, -.10], p < .001, though the effect size was negligible (R^2= 0.002, or .2% of the variance in diversity antagonism scores). This relationship remained significant when a random slope was added to the effect, b = -.15 (SE = .02), 95% CI [-.19, -.10], p < .001, and the Likelihood Ratio Test (LRT) indicated there was significant variation in the training-attitude link across forces (LRT = 9.97, p = .007)[2]. Because adding this random slope worsened model fit (increasing the BIC from 119,649 to 119,660) and more complex models should only be retained if the BIC drops (Raferty, 1995), we kept anti-bias training, as well as all other predictors, as fixed effects.[3]

In the adjusted model, the effect of anti-bias training remained significant, though still negligible, when accounting for the effects of other predictors (autonomy support, consequences, and demographic covariates), b = -.09 (SE = .02), 95% CI [-.13, -.05], p < .001, R^2= 0.001; see Table 1. Perceiving the force to provide autonomy support to reduce bias also linked to lower diversity antagonism, b = -.32 (SE = .01), 95% CI [-.35, -.30], p < .001, and this was a moderate effect, R^2= 0.05. Perceiving the force to provide autonomy support to reduce bias interacted with anti-bias training, b = -.11 (SE = .02), 95% CI [-.15, -.07], p < .001 (Figure 1), such that participating in anti-bias training when individuals perceived their forces to provide autonomy support to reduce bias related to lower diversity antagonism, b = -.21 (SE = .03), 95% CI [-.27, -

[2] As it is difficult to compare R^2 of random and fixed effects, we quantified this heterogeneity by running a random-effects meta-analysis. Approximately half of the variance in the training-attitude relation reflected differences in effect sizes across forces, I^2 = 45.33%, T^2 = .005, indicating substantial heterogeneity, $Q(42)$ = 73.19, p = .002.

[3] We ran models when slopes for all effects were set to random and found no difference in the strength or direction of any effects.

.15], $p < .001$, $R^2 = .002$, while anti-bias training did not relate to antagonism in forces that were low in autonomy support to reduce bias, $b = .03$ ($SE = .03$), 95% CI [-.03, .09], $p = .33$. However, this interaction term accounted for negligible variance in diversity antagonism, $R^2 = 0.001$.

Testing covariates, clear consequences for biased actions did not relate to attitudes, nor did it moderate the link with anti-bias training ps $< .05$, $R^2 = .000$. All demographic covariates significantly predicted diversity antagonism (see Table 1), though most of these explained a negligible amount of variance. Race and sex showed small effects, accounting for 3% and 1% of the variance, respectively.

Discussion

We tested the links of anti-bias training, perceived autonomy-supportive communication about bias, and their interaction on prejudiced attitudes across police forces in the U.K. Policing provides an important organizational context for these tests for three primary reasons: First, although prejudice is always consequential, the implications of prejudice are elevated in policing- for community members (Bor et al., 2018), policing employees (Brough & Frame, 2004), and the organization itself (Roberg et al., 2000). Second, policing invests heavily in anti-bias training but may not be gaining the needed benefits (Smith, 2015). And third, implementation can vary across separate units (i.e., forces) with distinct cultures. Our findings across 43 police forces within the U.K. identified a statistically significant relation between participating in anti-bias training and lower prejudiced attitudes. Yet the effect size ($R^2 = .002$, or $d = .09$) is quite small and consistent with those reported for follow-ups of laboratory interventions ranging from $d = .01$–$.12$ (Lai et al., 2016). There was considerable variability in the training-attitude link observed across forces, which may be due to error in responding,

characteristics of the individuals sampled, or ways in which anti-bias training was delivered. The very small effect size supports observations in healthcare settings that anti-bias training has a long way to go before reaching the ultimate translational goal: better outcomes for individuals from oppressed and marginalized groups (Hagiwara et al., 2020). The effect size observed here likely would not translate to meaningful observable relationships.

A more robust correlate of attitudes was perceiving autonomy support to reduce bias from the force; it showed a small to moderately-sized link to lower prejudice, $R^2 = .05$ or $d = .44$ – by far, the largest in the model. Perceiving autonomy support to reduce bias from the force also moderated the training-attitude link such that anti-bias training did not relate to prejudiced attitudes in the absence of perceived autonomy support from the force. However, even when individuals perceived high autonomy support from the force, the relationship between anti-bias training and attitudes was still quite small, $R^2 = .002$ or $d = .10$. Prior work shows that motivation may be a key factor in determining whether or not training is effective (Shamader et al., 2022; Wen, 2020), and these data modestly support this assertion. Similarly, Hagiwara and colleagues (2020) recommended promoting autonomous motivation to reduce bias as an integral part of training in healthcare, further underscoring the potential benefits of motivational approaches in the context of prejudice-reduction training.

Given the possibility that motivational principles can be applied to improve anti-bias training, future translational work is needed to identify and systematize concrete strategies for bias reduction in real-life settings (Ijzerman, et al., 2020). Further, these must be examined alongside solutions for police reform that are known to work (e.g., removing military equipment, Eberhardt, 2016; fatigue reduction, James, 2018; structural and institutional changes, Ray, 2020).

Methodological Considerations and Limitations

Two methods are typically used to evaluate anti-bias training: self-reported awareness of bias, or changes in implicit bias scores. While the two are highly correlated (Hahn et al., 2014), there is no evidence to assume implicit bias is a better measurement of attitudes. This study relied on a self-report measure of attitudes toward investing in diversity, an explicit measure with unknown links to actual behavior. An implicit bias measure may have produced a larger correlation with training than the explicit measure we used, though research has failed to find evidence that implicit bias scores map onto actual changes in behavior (Forscher et al., 2019). Observed and behavioral real-world measures of bias are needed to triangulate with self-report measures for robust translational research (Hagiwara et al., 2020). Ultimately, researchers should measure behavior over time since modest gains in attitude change observed in experimental work are not typically sustained (Lai et al., 2016).

The current findings should be interpreted with caution given the number of unmeasured variables (e.g., mandatory vs. voluntary training, time elapsed since exposure to training, specific content covered during training, and fidelity of training) and the correlational and retrospective nature of the data (as training could have happened days, months, or years ago). To further develop this work, it is important to obtain precise training participation data, and to gather data immediately after participation as well as at later time-points. Most crucially, examining the effectiveness of anti-bias training will require randomized designs, as there may be systematic differences in who participates in anti-bias training (e.g., self-selection bias, role differences). These variables cannot be accounted for without a closely controlled experimental design. Examining more nuanced effects of anti-bias training across diverse identities and demographic characteristics beyond the dichotomized variables we tested here is crucial for future work, including consideration of how the identities of the facilitator and attendees interact (Hornsey &

Eposo, 2009). To best understand how motivational factors such as autonomy support provision may impact training, it should be measured or manipulated within the trainings themselves, as well as asking participants to reflect about broader perceptions of the force (the organizational context). Such methods will allow researchers to test the independent contributions of culture and training, both of which are likely important for achieving attitudinal change (Tannenbaum & Yuki, 1992).

Conclusion

This study provided one of the first real-world tests of the question of whether anti-bias training is linked with biased attitudes in a policing context and examined training alongside an aspect of the broader organizational culture motivating prejudice reduction. Findings of this study showed a very small overall link between anti-bias training and lower prejudiced attitudes, as well as a moderate link between perceiving the force to communicate about bias in a way that promotes personal agency (versus motivating through shame and fear) relating to lower prejudice. Interestingly, training only appeared to relate to lower prejudice when the force communicated about bias with autonomy support (though the effect was still quite small); there was no relation in the absence of an autonomy-supportive force climate. Taken together, this study suggests that while research should continue to investigate the effectiveness of antibias training and ways to enhance it, it is equally important to study how the broader organizational climate can motivate sustained prejudice reduction in their employees.

References

Al-Khouja, M., Graham, L., Weinstein, N., & Zheng, Y. (2020). How autonomy support and ethical value alignment influences attitudes towards diversity in English police. *Journal of Moral Education*, *49*(3), 365-380.

Applebaum, B. (2019). Remediating campus climate: Implicit bias training is not enough. *Studies in Philosophy and Education*, *38*(2), 129-141.

Atewologun, D., Cornish, T., & Tresh, F. (2018). Anti-bias training: An assessment of the evidence for effectiveness. *Equality and Human Rights Commission Research Report Series*.https://cstar.ie/equality/t4media/ub_an_assessment_of_evidence_for_effectiveness.pdf

Bezrukova, K., Spell, C. S., Perry, J. L., & Jehn, K. A. (2016). A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychological Bulletin*, *142*(11), 1227-1274.

Bliese, P. D., Maltarich, M. A., & Hendricks, J. L. (2018). Back to basics with mixed-effects models: Nine take-away points. *Journal of Business and Psychology*, *33*(1), 1-23.

Bor, J., Venkataramani, A. S., Williams, D. R., & Tsai, A. C. (2018). Police killings and their spillover effects on the mental health of black Americans: A population-based, quasi-experimental study. *The Lancet*, *392*(10144), 302-310. 10.1016/S0140-6736(18)31130-9

Brough, P., & Frame, R. (2004). Predicting police job satisfaction and turnover intentions: The role of social support and police organisational variables. *New Zealand Journal of Psychology*, *33*(1), 8-18.

Carter, E. R., Onyeador, I. N., & Lewis, N. A., Jr. (2020). Developing & delivering effective anti-bias: Challenges & recommendations. *Behavioral Science & Policy*, 6(1), 57–70.

Chan, J. B. (1997). *Changing police culture: Policing in a multicultural society*. Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooper, H. L., & Fullilove, M. T. (2020). *From Enforcers to Guardians: A Public Health Primer on Ending Police Violence*. Johns Hopkins University Press.

Eberhardt, J. L. (2016). *Strategies for change: Research initiatives and recommendations to improve police-community relations in Oakland, Calif.* Stanford University, SPARQ: Social Psychological Answers to Real-world Questions.

Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., & Schabenberger, O. (2008). An R2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine*, *27*(29), 6137-6157.

FitzGerald, C., Martin, A., Berner, D., & Hurst, S. (2019). Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: A systematic review. *BMC Psychology, 7*(1), 29.

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology, 117,* 522–59.

Hagiwara, N., Kron, F. W., Scerbo, M. W., Watson, G. S. (2020). A call for grounding implicit bias training in clinical and translational frameworks. *The Lancet*.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General, 143*(3), 1369–1392.

Hannah, S. D., Carpenter-Song, E. (2013). Patrolling your blind spots: Introspection and public catharsis in a medical school faculty development course to reduce unconscious bias in medicine. *Culture, Medicine, and Psychiatry, 37,* 314–339.

HMICFRS; Her Majesty's Inspectorate of Constabulary. (2017). *PEEL: Police Legitimacy 2017: A National Overview,* London.

Hornsey, M. J., & Esposo, S. (2009). Resistance to group criticism and recommendations for change: Lessons from the intergroup sensitivity effect. *Social and Personality Psychology Compass*, *3*(3), 275-291.

IJzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., ... & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*, *4*(11), 1092-1094.

James, L. (2018). The stability of implicit racial bias in police officers. *Police Quarterly*, *21*(1), 30-52.

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review*, *71*(4), 589-617.

Kirkland, R. & Bohnet, I. (2017, April 7). Focusing on what works for workplace diversity. *McKinsey & Company*. https://www.mckinsey.com/featured-insights/gender-equality/focusing-on-what-works-for-workplace-diversity.

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., ... & Simon, S. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*(8), 1001-1016.

Lawrence, F. M. (2009). *Punishing hate*. Harvard University Press.

Legault, L., Gutsell, J. N., & Inzlicht, M. (2011). Ironic effects of antiprejudice messages: How motivational interventions can reduce (but also increase) prejudice. *Psychological Science*, *22*(12), 1472-1477.

Lewis Jr, N., & Wai, J. (2020). Communicating what we know, and what isn't so: Science communication in psychology. PsyArXiv.

Noon, M. (2018). Pointless diversity training: Unconscious bias, new racism and agency. *Work, Employment and Society*, *32*(1), 198-209.

Oberai, H. & Anand, I. M. (2018). Unconscious bias: Thinking without thinking. *Human Resource Management International Digest, 26*(6), 14-17.

Onyeador, I. N., Hudson, S. K. T., & Lewis Jr, N. A. (2021). Moving beyond implicit bias training: Policy insights for increasing organizational diversity. *Policy Insights from the Behavioral and Brain Sciences*, *8*(1), 19-26.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171-192.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111–163.

Ray, R. (2020, Aug. 25). How can we enhance police accountability in the United States? *The Brookings Institution.* https://www.brookings.edu/policy2020/votervital/how-can-we-enhance-police-accountability-in-the-united-states/

Roberg, R. R., Crank, J. P., & Kuykendall, J. L. (2000). *Police & society*. Los Angeles, CA: Roxbury.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications. ISBN: 1462528767

Schmader, T., Dennehy, T. C., & Baron, A. S. (2022). Why antibias interventions (need not) fail. *Perspectives on Psychological Science*, Online first https://doi.org/10.1177/17456916211057565.

Smith, R. J. (2015). Reducing racially disparate policing outcomes: Is implicit bias training the answer? *Harvard Law Review*, *37*, 295.

Souhami, A. (2014). Institutional racism and police reform: An empirical critique. *Policing and Society*, *24*(1), 1-21.

Tannenbaum, S. I., & Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology*, *43*(1), 399-441.

Tyler, T. R., & Huo, Y. (2002). *Trust in the law: Encouraging public cooperation with the police and courts*. Russell Sage Foundation.

Wen, T. (2020, Aug 28). Police forces and many other institutions are turning to implicit bias training to help their staff recognise when they are relying upon racist assumptions and stereotypes. But does it really work? *The British Broadcasting Corporation.* https://www.bbc.com/future/article/20200827-is-it-possible-to-rid-police-officers-of-bias

Weinstein, N., Legate, N., Graham, L., Zheng, Y., Plater, M., Al-Khouja, M., Moller, A. C. (2022). The role of perceived autonomy-supportive communications for motivating prejudice reduction and avoiding defiant backlash within the police force workplace. Unpublished manuscript.

Worden, R. E., McLean, S. J., Engel, R. S., Cochran, H., Corsaro, N., Reynolds, D., Najdowski, C. J., & Isaza, G. T. (2020). *The impacts of implicit bias awareness training in the NYPD*. IACP/UC Center for Police Research and Policy & John F. Finn Institute for Public Safety.

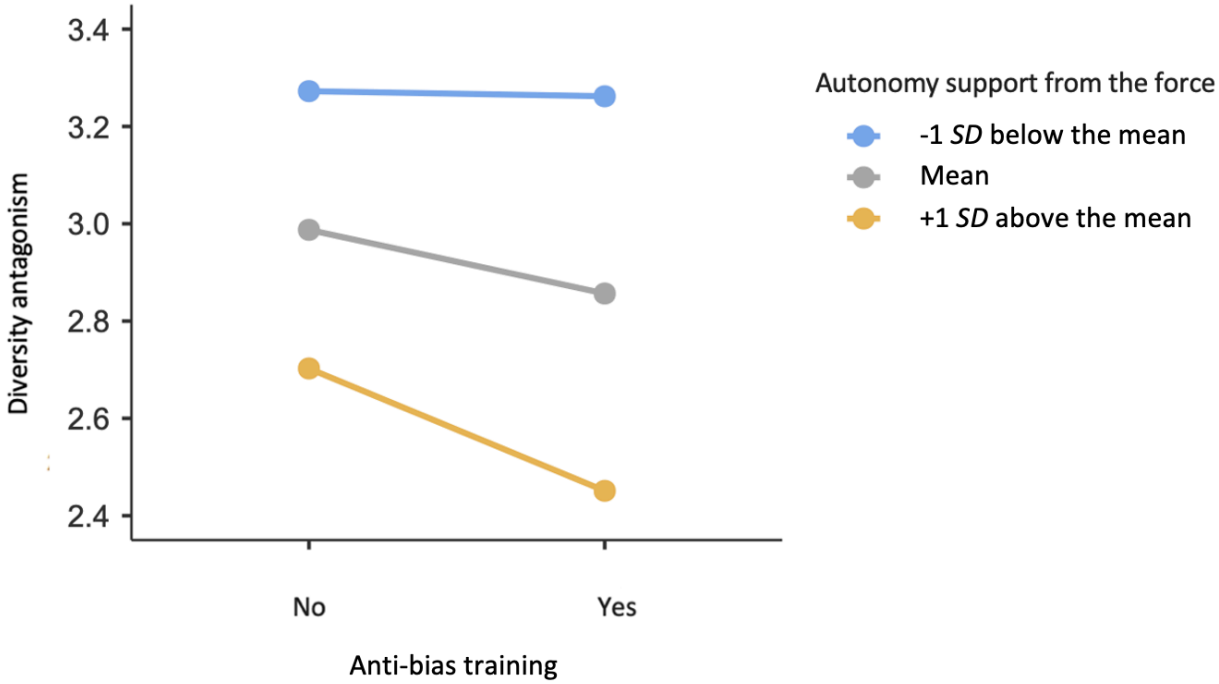
Table 1.

Results of unadjusted and adjusted linear mixed models predicting diversity antagonism

	B (*SE*)	95% CI	R^2
Unadjusted			
Anti-bias training	**-.14 (.02)**	**-.17, -.10**	**.002**
Adjusted			
Anti-bias training	**-.09 (.02)**	**-.13, -.05**	**.001**
Autonomy support	**-.32 (.01)**	**-.35, -.30**	**.05**
Clear consequences	**.00 (.01)**	**-.02, .02**	**.000**
Age	**.09 (.01)**	**.07, .12**	**.003**
Sex	**-.30 (.02)**	**-.34, -.26**	**.01**
Race	**1.05 (.05)**	**.95, 1.14**	**.03**
Tenure	**.07 (.01)**	**.05, .09**	**.002**
Role	**-.20 (.02)**	**-.24, -.15**	**.004**
Training X autonomy support	**-.11 (.02)**	**-.16, -.07**	**.001**
Training X clear consequences	.02 (.02)	-.02, .05	.000
Training X age	-.03 (.03)	-.08, .02	.000
Training X sex	-.04 (.04)	-.12, .05	.000
Training X race	.09 (.09)	-.09, .28	.000
Training X tenure	**-.06 (.02)**	**-.11, -.02**	**.000**
Training X role	.03 (.04)	-.06, .12	.000

Note. *B*s represent the unstandardized regression coefficients and *SE*s are their standard errors; Anti-bias training is coded as 1 yes, 0 no; Sex is coded as 0 = male, 1 = female; Role is coded as 1 = officers, 2 = staff; Race is coded as 1 = BAME (Black, Asian and minority ethnic), 2 = White; All other variables were treated as continuous in analyses; Bolded effects indicate statistical significance at $p < .05$

Figure 1. Interaction of anti-bias training and the force communicating about bias in autonomy-supportive ways predicting diversity antagonism



Note. The moderator, the force communicating about bias in autonomy-supportive ways, is split at one standard deviation above the mean, the mean, and one standard deviation below the mean.