

The use of measures of explicit and implicit bias for predicting discrimination

Authors: Calvin K. Lai & John F. Dovidio

Date: July 23, 2021

Measuring prejudice and bias

Selecting peace officers based on their having low levels of pre-existing prejudice may be particularly important for preventing **discrimination** (i.e., unfair treatment or behaviors directed toward people on the basis of their group membership) in policing. In large part, this is because intervention or training efforts to reduce prejudice in an enduring way after hiring have shown only limited effectiveness. We define **prejudice (or bias)** as an attitude (or response) toward a group and its members that creates or maintains status differences between groups (Dovidio et al., 2010). Specifically, interventions and training have relatively small prejudice-reduction effects in large samples (average Cohen's $d = .19$), and very little of this research has been conducted in field settings or examined the persistence of prejudice reduction over long time periods (Paluck et al., 2021). Research on the impact specifically of diversity trainings, which educate people about racial bias, also shows limited success. While diversity trainings are consistently effective for changing attitudes and beliefs about diversity (Bezrukova et al., 2016), research in naturalistic (“real-world”) settings has a mixed record for reducing actual discriminatory behavior (e.g., Chang et al., 2019; Worden et al., 2020).

Moreover, prejudice and bias (involving responses to members of another group *relative to* members of one’s own group) can be **explicit** in the sense that they are attitudes or beliefs about a group and its members that people know they hold and are willing to express. Prejudice or bias can also be **implicit**, in the sense that they may often be activated without intention or awareness. Implicit biases tend to be relatively less controllable, less conscious, and retrieved more quickly and efficiently from memory. Research on reducing implicit bias specifically has not identified interventions that are consistently effective for creating long-term change (Lai et al., 2016). Even when interventions are effective for reducing implicit biases in the short-term, those reductions do not guarantee a corresponding reduction in discrimination (Forscher, Lai, et al., 2019). The limited available evidence for the use of training to reduce prejudice, bias, and discrimination highlights the importance of selecting peace officers who will treat all members of the community fairly and equitably. The peace officer selection process can be informed by studies in other contexts demonstrating relationships between measures of bias and discriminatory action.

The development of implicit measures of prejudice and bias (representing differences in the favorability of the associations for one group over another) was stimulated by concerns that people were systematically misrepresenting their “true” attitudes on explicit measures because of concerns about being socially “correct” or wanting to conform to expectations in a particular setting (i.e., social desirability bias). Researchers were also noticing that beliefs about or

perceptions of groups and their members were not necessarily deliberate or controlled (e.g., Devine, 1989). As noted earlier, whereas explicit measures ask people to directly report their feelings or beliefs about a group, measures of implicit bias reveal feelings or beliefs through performance on a task that does not appear to be related directly to prejudice. The **Implicit Association Test (IAT)**, which is the most commonly used implicit measure, assesses response times to positive and negative words that are presented in combination with specified social groups (e.g., Black and White people, gay and straight people). The IAT draws on an extensive literature demonstrating that faster response times reflect stronger mental associations (Greenwald et al., 1998). Despite some critiques (e.g., Blanton et al., 2015), the IAT is well-validated (Greenwald et al., 2021) and has largely superseded older indirect measures of attitude such as sentence completion tasks or projective tasks like the Thematic Apperception Test. Measures of implicit bias, such as the IAT, reveal that evaluations of groups and their members can be automatically activated, suggesting that prejudicial attitudes need not involve intention or personal endorsement. Although people may sometimes not be aware that they hold implicit biases, implicit biases are not necessarily unconscious. Implicit biases may sometimes be characterized by a lack of control (e.g., an intrusive thought) or by their speed or efficiency (e.g., a split-second reaction). For example, a peace officer's racialized feeling of suspicion about a Black man in a predominantly White suburb may be a bias that the officer is aware of consciously, but it may nonetheless be implicit due to the difficulty of ignoring that racialized feeling and the speed at which it comes to mind. Measures of explicit and implicit intergroup bias are positively related, but usually only moderately so.

In addition to measures of prejudice toward specific groups, measures of general ideologies about intergroup relations, such as **Social Dominance Orientation (SDO)** (Sidanius et al., 2017) and **Right-Wing Authoritarianism (RWA)** (Altemeyer, 1998) can reflect biases toward a range of other groups without delineating a specific group in the scale's items. People in the US and in numerous other countries who score higher in SDO or RWA tend to be more prejudiced in the attitudes they hold and more discriminatory in their behavior toward traditionally marginalized groups in their society (Duckitt & Sibley, 2017).

Relationships between bias and intergroup behavior

Preferences for collective security, control, stability, and order (i.e., RWA) and for hierarchy, power, and dominance (i.e., SDO) tend to increase in social contexts involving threat or competitiveness from or toward other groups, respectively (Duckitt & Sibley, 2017). There also is evidence that individuals seek roles in an institution expected to be compatible with their SDO levels (Sidanius et al., 1994), and this finding may explain why some studies report that police officers, and particularly White police officers, tend to score high on measures of SDO (Sidanius et al., 1994; replicated in data provided by Xu et al., 2014) and show a range of prejudices that increased with training and experience in policing (Gatto et al., 2010; see also Pratto et al., 2006).

Explicit and implicit bias predispose people to act with a negative orientation toward a group and individual members of a group. Across a broad range of situations and behaviors, measures of explicit and implicit bias both reliably (but only weakly or moderately) predict discrimination, and three of the four meta-analyses studying the topic indicate that they do so

uniquely from each other (Cameron et al., 2012; Greenwald et al., 2009; Kurdi et al., 2019; Oswald et al., 2013). Most of the relevant research has been in controlled laboratory settings. However, field studies in naturalistic work settings have demonstrated consistent and continuing racial discrimination in hiring (Quillian et al., 2017) that is predicted by individual levels of prejudice (e.g., Glover et al., 2017; Rooth, 2010). That said, intergroup behaviors are determined by multiple forces, not only by personal prejudice and bias but also by situational constraints or expectations, perceptions about how one should behave, the amount of investment required to pursue alternative courses of action, and cost-benefit analyses for different behaviors. As discrimination is multiply determined (e.g., affected by normative pressures, context, and opportunity), both explicit-discrimination and implicit-discrimination relationships tend to be small to moderate.

Moderators of the relationship between bias and behavior

There is no single unconditional relationship between bias and behavior (Gawronski, 2019). The relative validity of explicit and implicit measures for predicting behavior depends on the topic, the context in which the behavior occurs, and the type of behavior being examined (Dovidio et al., 2009; Lai & Wilson, 2020). For example, the **Motivation and Opportunity as DEterminants of attitude-to-behavior processes (MODE) model** suggests that explicit measures would be better at predicting deliberate responses than implicit measures, as deliberate responses are more subject to conscious control and awareness. In contrast, implicit measures would better predict spontaneous behaviors (Fazio & Olson, 2003, 2014). The pattern of findings reported by Kurdi et al.'s (2019) meta-analysis of 217 research reports is partially consistent with the MODE model. Consistent with predictions for explicit measures, Kurdi et al. found that explicit measures better predicted intergroup behavior in contexts in which participants had greater control of their behavior and were more aware that discrimination was being measured. However, inconsistent with predictions derived for implicit measures, Kurdi et al. found that implicit measures systematically predicted deliberative and spontaneous behaviors to a comparable degree.

Thus, implicit bias has a general influence on discriminatory behavior, unrelated to the spontaneity of the action. This may be because even when considerable deliberation is involved implicit bias may guide what people notice and how they recall and interpret information at a relatively early stage of processing. For example, when police officers are placed in situations in which crime is emphasized, they attend more quickly to faces of Black men than of White men, and be more likely to falsely identify Black men in lineups (Eberhardt et al., 2004). Police officers also display less respect when interacting with Black than White motorists in traffic stops (Voigt et al., 2017). The spontaneity of these actions suggests the influence of implicit processes, but direct research on the role of individual differences in implicit bias in such behaviors comes primarily from outside of policing. For example, research in medicine finds that physicians with stronger implicit racial biases respond in a less friendly and respectful way when interacting with Black patients (Penner et al., 2014).

In addition to differences in how well explicit and implicit biases predict various outcomes, the degree to which explicit and implicit biases align for individuals also matter for behavioral prediction. Measures of explicit and implicit biases are weaker predictors of

intergroup behavior when they are less strongly related with each other (Cameron et al., 2012; Greenwald et al., 2009; Greenwald & Lai, 2020; Kurdi et al., 2019). This may reflect a form of ambivalence. When biases are aligned, explicit and implicit biases may mutually reinforce each other to guide behavior. When biases are misaligned, explicit and implicit biases may "compete" in guiding behavior. This competition would then reduce the causal influence of both explicit and implicit biases.

Explicit measures, personnel selection, and policing behavior

We also conducted a literature review to examine whether five popular measures of explicit biases had been used in research on personnel selection (i.e., Social Dominance Orientation, Right-Wing Authoritarianism, **Ambivalent Sexism Inventory**, **Attitudes toward Lesbians and Gay Men**, **Bias Awareness**). We did not find evidence that any of these measures had been used or adapted for selecting personnel, in policing or elsewhere.

In addition to the literature on the use of the five common measures of bias in personnel selection, we reviewed literature on the use of these measures for predicting the beliefs and actions of members of law enforcement. Only a couple of the measures had been used with law enforcement populations (e.g., Bernstein et al., 2003; Sidanius et al., 1994). Most of the researchers using these measures only examined the average levels of bias among police officers or trainees relative to other samples; they generally did not investigate whether the measures predicted beliefs or behavior relevant to policing. There were, though, two exceptions. One exception was research by Swencionis et al. (2021) who found that White officers (but not non-White officers) with higher levels of Social Dominance Orientation had a relatively larger number of use of force incidents). The demographic status (e.g., race or ethnicity) of those residents were not examined in that study for discriminatory application of force, however. The other exception found that officers who endorsed Hostile or Benevolent Sexism were more likely to endorse myths about rape (e.g., beliefs that rape happens suddenly at night with aggression and clear evidence of resistance) that would interfere with effective policing (Murphy & Hine, 2019).

Finally, we examined the extent to which measures of bias were related to discrimination (see prior sections). Research on the relationship between bias and discrimination tends to focus on undergraduate student samples or discrimination in non-law enforcement workplace settings, suggesting that more research is needed in law enforcement, personnel selection, and screening.

Implicit measures and discrimination in policing

Researchers studying implicit measures have not directly examined personnel selection or discrimination in actual policing behavior. Instead, the research on implicit measures and discrimination in policing has often been focused on simulations of decisions to shoot armed or unarmed men. In the most common form of these simulations, participants engage in a computer simulation in which they observe images of White and Black men in everyday places like parks or city sidewalks (Correll et al., 2002; Payne & Correll, 2020). Some of the men are armed with guns, while others are unarmed and carrying mundane objects like wallets or cell phones. Participants are instructed to press a button to "shoot" if the man is holding a gun or another

button to "don't shoot" if the man is not holding a gun. In this simulation, civilians tend to mistakenly shoot unarmed Black men at higher rates than unarmed White men. Civilians also tend to react faster to unarmed Black men than unarmed White men. Follow-up research on the intersection between race and gender has found that Black women are not more likely to be mistakenly shot than White women or White men in these simulations, suggesting that Black men are uniquely associated with violence and danger (Plant et al., 2011).

Police officers' racial biases in decisions to shoot are more variable, and these effects depend upon the life experiences of the police officers. On average, officers with policing experience do not tend to show a racial bias in decisions to shoot (Correll et al., 2007). This average obscures variability in officers' biases. Officers who have less childhood contact with Black people or are assigned to gang and street-crime units that regularly deal with minority gang members tend to show higher levels of racial shooting bias (Sadler et al., 2012; Sim et al., 2013). New police recruits also tend to show evidence of a racial shooting bias, suggesting that certain on-the-job experiences may be important for mitigating racial shooting biases (Ma et al., 2013). Finally, the situation matters: Officers who are more tired show stronger racial bias in decisions to shoot, whereas officers who are more alert show less racial bias (Ma et al., 2013). Administrative policing records from Chicago also provide evidence supporting the critical role of life experience in police use of force (Ba et al., 2021). Hispanic and Black officers in Chicago are generally less likely to use force than their White counterparts, and they are especially less likely to use force when policing majority-Black areas and Black community members.

Other factors related to explicit and implicit bias

When there is clear and direct evidence of explicit or implicit bias (e.g., via a measure of Social Dominance Orientation), other factors are not relevant for assessing the bias of an officer candidate. However, when direct evidence of bias is ambiguous or weak, other factors that are related to explicit or implicit bias become more important in assessment. In addition, even when there is some direct evidence of prejudice or bias, these factors can override these predispositions, reducing the likelihood that people will act in a discriminatory way. Here, we describe some of these relevant factors.

First, experiences of **intergroup contact** (i.e., interactions with people of other groups) has been correlated with a range of positive intergroup outcomes such as reduced prejudice, stereotyping, and discrimination (Pettigrew, 1998). Second, chronic **motivations to respond without prejudice** have been linked to reduced discrimination, which highlights the importance of self-control in preventing discrimination (Butz & Plant, 2009). In particular, intrinsic motivations to be unprejudiced out of a commitment to egalitarian values are particularly potent for combatting expressions of bias. In comparison, extrinsic motivations to be unprejudiced due to concerns about reputation tend to have a more mixed track record for combatting bias. Third, **social norms about prejudice** can powerfully reduce the expression of prejudice in daily life (Crandall et al., 2002). Perceiving that being prejudiced or acting in discriminatory ways violates prevailing social norms motivates individuals to regulate their biases. Experience in regulating these responses can produce internalized standards of fairness (e.g., "To be a good police officer, I need to be unprejudiced") and arouse concerns about belonging (e.g., "To fit in with the other officers, I need to be unprejudiced"). Finally, higher levels of **executive function**, or cognitive

control over one's own behavior, has been linked to reduced expressions of implicit bias in behavior (Ito et al., 2015).

Limitations to the measurement of explicit and implicit bias in personnel selection contexts

Conceptual Limitations

Based on our broader understanding of the research, we identify limitations of using explicit or implicit measures of bias as determining factors in personnel selection, except in their extreme expressions. First, many measures of explicit bias often contend with tendencies to engage in **socially desirable responding** (Schuman et al., 1997). This means that many participants may censor themselves and report lower levels of prejudice than what they really believe. Under situations with high evaluative stakes like personnel selection, these motivations to self-censor may be further amplified (Lerner & Tetlock, 1999). That could mean that higher levels of self-reported bias may be capturing honesty or a lack of awareness of norms about expressing prejudice rather than levels of prejudice alone.

Second, measures of implicit bias are currently not diagnostic of individuals due to their relatively low **test-retest reliability**. In comparison to measures of explicit bias which have an average test-retest reliability of .75 (Gawronski et al., 2017), the Implicit Association Test has an average test-retest reliability of .49 (Lai & Wilson, 2021). This means that a single administration of the Implicit Association Test will predict only 25% of the performance in a follow-up Implicit Association Test (on average). There are ongoing efforts and possible approaches to increase the **reliability** of implicit measures for diagnostic use, but such approaches have not yet been published in peer-reviewed scientific journals (for a general guide to best practices in Implicit Association Test measurement, see Greenwald et al., 2021). In addition to test-retest reliability, there are also other potential issues for screening such as the possibility of deliberate faking (Fiedler & Bluemke, 2005), the interpretation of scores as a measure of absolute preference (Blanton et al., 2015), and noisiness in measurement due to contamination from non-prejudice-related processes (Calanchini & Sherman, 2013).

Third, the expression of explicit and implicit bias is often better understood through the characteristics of the social environment rather than an individual's chronic level of prejudice. For police officers, situational factors like time pressure, the experience of identity threat, and sheer inexperience on the job are all linked to the expression of racial bias (Swencionis & Goff, 2017). Ambiguity and personal discretion are particularly relevant factors in policing. In their day-to-day work, police officers must contend with high degrees of ambiguity (e.g., determining who fits a suspect description or who poses a potential threat) and have high discretion (e.g., deciding whether to search in a traffic stop). This is important because the propensity to discriminate is more likely when there is high ambiguity in how to make a decision and decision-makers have high discretion to act on bias. Consistent with this proposition, policing records show evidence of biased discretion during traffic stops and police interactions in California. Officers chose to search individuals that were perceived to be Black at over twice the rates of individuals that were perceived to be White. However, the probability of finding contraband

during the search (i.e., the discovery rate) was higher for individuals that were perceived to be White (Racial and Identity Profiling Advisory Board, 2021), indicating evidence of racial discrimination.

Social psychological experiments further demonstrate the critical role for ambiguity and discretion in discrimination. For example, White individuals evaluated Black and White job candidates that varied in the strength of their qualifications (Dovidio & Gaertner, 2000). When the candidates' qualifications were clearly strong or clearly weak, people did not racially discriminate. Racial discrimination was only evident when the qualifications were more ambiguous (i.e., a mix of qualities that fit and did not fit the position).

That said, an integrated perspective that considers the interaction between the person and the situation may sometimes be relevant. For example, Son Hing et al. (2008) found that people who had stronger implicit bias against Asians less strongly supported the hiring of a moderately qualified Asian applicant. In contrast, implicit bias against Asians did not affect decisions when the Asian applicant was highly qualified, presumably because there was less ambiguity and less discretion. Thus, even when the effects of implicit (or explicit) bias may be limited on average, considering how attitudes influence behaviors in different types of situations (a person x situation perspective) can yield conceptually meaningful and practically important findings.

Limitations of using measures of general social skills to predict discrimination

There are a broad range of social skills or personality traits that are important for social functioning generally (e.g., empathy, theory of mind, perspective-taking, a tendency to think before speaking, Agreeableness, Openness to Experience). These social skills tend to have broad positive influences for how people treat others in everyday life. However, the effects of these broad social skills on prejudice and discriminatory behavior (i.e., *group-based differences* in how we treat others) are often weaker and more contingent. For example, the relationship between measures of dispositional empathy and explicit prejudice is weakly negative overall (McFarland, 2010). In practice, however, dispositional empathy can manifest in more care toward *one's own group* (e.g., supporting war in another country in defense of one's nation) rather than more care toward other groups (e.g., opposing war due to the consequences for civilians in that other country; Bloom, 2017; Peak et al., 2016). To address this discrepancy, some interventions have been developed to tune these social skills for addressing prejudice, such as efforts to increase empathy or perspective-taking toward members of marginalized groups (Shih et al., 2009; Stephan & Finlay, 1999).

Summary and Practical Takeaways

Given the limited evidence of the effectiveness of efforts to train peace officers to become less biased on the job, it is important to instead consider how psychological evaluators can play a role in screening out candidates who hold explicit or implicit biases about race or ethnicity, gender, nationality, religion, disability, or sexual orientation. Our literature review finds that there is a lack of direct scientific evidence about how to effectively and reliably screen for explicit or implicit bias in a personnel selection context. However, there are many studies

(primarily conducted in research settings) which led us to the following conclusions that are relevant for screening peace officer candidates effectively:

1. Discrimination is multi-determined, involving individuals' explicit and implicit biases, the social situation, and the opportunities and consequences for expressing bias. As a result, measures of explicit and implicit bias will only weakly or moderately predict discrimination.
2. We found no studies on the measurement of explicit and implicit bias in personnel selection contexts. However, due to high demand characteristics in these settings, measures of explicit bias particularly (Detrick & Chibnall, 2014) and implicit bias to a lesser degree may be especially subject to faking or other forms of socially desirable responding. Of the explicit measures we reviewed, the Social Dominance Orientation, Ambivalent Sexism Inventory Scale, and beliefs about rape myths showed the strongest evidence for predicting biased policing.
3. When direct evidence of bias is unavailable, ambiguous, or weak, it may be useful to consider, in assessments and/or interviews, related factors such as favorable intergroup contact, motivations to respond without prejudice, perceptions of social norms about prejudice, and executive function. These factors also generally contribute to more equitable behavior and fair treatment of others and can mitigate tendencies to act in discriminatory ways even when some evidence of bias is detected.
4. More distantly related social skills like a general tendency toward empathy or perspective-taking are consistently related to generally more positive treatment of others. However, because these may be applied more readily to members of one's own group rather than to members of other groups, they are not consistently linked to less group-based differences in treatment (i.e., discrimination).

Glossary

Ambivalent Sexism Inventory: Measure of sexism that encompasses hostile sexism (subjectively negative evaluations of women) and benevolent sexism (subjectively positive evaluations of women that reinforce restrictive traditional gender roles)

Attitudes Toward Lesbians and Gay Men: A direct measure of prejudice toward lesbian women and gay men

Bias: A response toward a group and its members, relative to one's own group, that creates or maintains status differences between groups

Bias Awareness: Measure of awareness and concern about racial bias

Discrimination: Unfair treatment or behaviors directed toward people on the basis of their actual or perceived group membership

Executive function: Cognitive processes that are necessary for controlling behavior

Explicit bias (or explicit prejudice): An attitude or belief (e.g., stereotype) about a group and its members that people know they hold and are willing to express. Explicit biases are more controllable, more conscious, and/or slower or less efficient to retrieve from memory than implicit bias. Explicit biases are typically assessed directly with self-report measures.

Implicit bias: Feelings or beliefs about a group and its members that are less controllable, less conscious, and/or faster or more efficient to retrieve from memory than explicit bias. Implicit biases may often be activated without intention or awareness. Implicit biases are typically assessed indirectly through performance on an ostensibly unrelated task.

Implicit Association Test (IAT): The most commonly used measure of implicit biases that assesses bias by measuring response times in categorizing social groups (e.g., Black and White people, gay and straight people) with attributes (e.g., good or bad)

Intergroup contact: Interactions between members of different groups

Motivation and Opportunity as *DE*terminants of attitude-to-behavior processes (MODE) model: Theory that predicts that implicit attitude measures would better predict spontaneous behaviors and explicit attitude measures better predict deliberative behaviors

Motivations to respond without prejudice: Motivation to act without prejudice, which can distinguish between internal motivation (to be egalitarian) and external motivation (to appear unprejudiced to others)

Norms about prejudice: Perceptions of whether expressing prejudice is normal or accepted within a social situation

Prejudice: An attitude toward a group and its members that creates or maintains status differences between groups

Reliability. A property of a psychological instrument or measurement reflecting its consistency in assessing the quality of interest

Right-Wing Authoritarianism (RWA): A trait encompassing submissiveness to authority, aggression in the name of authority, and a tendency toward conformity

Social Dominance Orientation (SDO): A trait encompassing support for social hierarchy and a desire for some groups to dominate other groups

Socially desirable responding: Answering surveys in ways that respondents believe will lead others to view them favorably

Stereotype: A belief about characteristics shared by a group of people

Test-retest reliability: A measure of consistency in a psychological assessment across time

References

- Altemeyer, B. (1998). The other “authoritarian personality.” In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 30, pp. 47–92). Academic Press.
[https://doi.org/10.1016/S0065-2601\(08\)60382-2](https://doi.org/10.1016/S0065-2601(08)60382-2)
- Ba, B. A., Knox, D., Mummolo, J., & Rivera, R. (2021). The role of officer race and gender in police-civilian interactions in Chicago. *Science*, 371(6530), 696–702.
<https://doi.org/10.1126/science.abd8694>
- Bernstein, M., Kostelac, C., & Gaarder, E. (2003). Understanding “heterosexism:” Applying theories of racial prejudice to homophobia using data from a southwestern police department. *Race, Gender & Class*, 10(4), 54–74.
- 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. <https://doi.org/10.1037/bul0000067>
- Blanton, H., Jaccard, J., Strauts, E., Mitchell, G., & Tetlock, P. E. (2015). Toward a meaningful metric of implicit prejudice. *Journal of Applied Psychology*, 100(5), 1468–1481.
<https://doi.org/10.1037/a0038379>
- Bloom, P. (2017). *Against empathy: The case for rational compassion*. Random House.
- Butz, D. A., & Plant, E. A. (2009). Prejudice control and interracial relations: The role of motivation to respond without prejudice. *Journal of Personality*, 77(5), 1311–1342.
<https://doi.org/10.1111/j.1467-6494.2009.00583.x>
- Calanchini, J., & Sherman, J. W. (2013). Implicit attitudes reflect associative, non-associative, and non-attitudinal processes. *Social and Personality Psychology Compass*, 7(9), 654–667. <https://doi.org/10.1111/spc3.12053>
- Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, 16(4), 330–350.
<https://doi.org/10.1177/1088868312440047>
- Chang, E. H., Milkman, K. L., Gromet, D. M., Rebele, R. W., Massey, C., Duckworth, A. L., & Grant, A. M. (2019). The mixed effects of online diversity training. *Proceedings of the National Academy of Sciences*, 116(16), 7778–7783.
<https://doi.org/10.1073/pnas.1816076116>
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer’s dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83(6), 1314–1329. <https://doi.org/10.1037/0022-3514.83.6.1314>
- Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keese, T. (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. <https://doi.org/10.1037/0022-3514.92.6.1006>

- Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, 82(3), 359–378.
- Detrick, P., & Chibnall, J. T. (2014). Underreporting on the MMPI–2–RF in a high-demand police officer selection context: An illustration. *Psychological Assessment*, 26(3), 1044–1049. <https://doi.org/10.1037/pas0000013>
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56(1), 5–18.
- Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, 11(4), 315–319. <https://doi.org/10.1111/1467-9280.00262>
- Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). *Handbook of prejudice, stereotyping, and discrimination*. Sage.
- Dovidio, J. F., Kawakami, K., Smoak, N., & Gaertner, S. L. (2009). The roles of implicit and explicit processes in contemporary prejudice. *Attitudes: Insights from the New Implicit Measures*, 165–192.
- Duckitt, J., & Sibley, C. G. (2017). The dual process motivational model of ideology and prejudice. In C. G. Sibley & F. K. Barlow (Eds.), *The Cambridge handbook of the psychology of prejudice* (pp. 188–221). Cambridge University Press. <https://doi.org/10.1017/9781316161579.009>
- Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing Black: Race, crime, and visual processing. *Journal of Personality and Social Psychology*, 87(6), 876–893. <https://doi.org/10.1037/0022-3514.87.6.876>
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 54(1), 297–327. <https://doi.org/10.1146/annurev.psych.54.101601.145225>
- Fazio, R. H., & Olson, M. A. (2014). The MODE Model: Attitude-behavior processes as a function of motivation and opportunity. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 155–171). Guilford Press.
- Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology*, 27(4), 307–316. https://doi.org/10.1207/s15324834basps2704_3
- 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(3), 522–559. <https://doi.org/10.1037/pspa0000160>
- Gatto, J., Dambrun, M., Kerbrat, C., & Oliveira, P. D. (2010). Prejudice in the police: On the processes underlying the effects of selection and group socialisation. *European Journal of Social Psychology*, 40(2), 252–269. <https://doi.org/10.1002/ejsp.617>
- Gawronski, B. (2019). Six lessons for a cogent science of implicit bias and its criticism. *Perspectives on Psychological Science*, 14(4), 574–595. <https://doi.org/10.1177/1745691619826015>

- Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin*, 43(3), 300–312. <https://doi.org/10.1177/0146167216684131>
- Glover, D., Pallais, A., & Pariente, W. (2017). Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores. *The Quarterly Journal of Economics*, 132(3), 1219–1260. <https://doi.org/10.1093/qje/qjx006>
- Greenwald, A. G., Brendl, M., Cai, H., Cvencek, D., Dovidio, J. F., Friesse, M., Hahn, A., Hehman, E., Hofmann, W., Hughes, S., Hussey, I., Jordan, C. H., Jost, J., Kirby, T. A., Lai, C. K., Lang, J. W. B., Lindgren, K. P., Maison, D., Ostafin, B., ... Wiers, R. (2021). The Implicit Association Test at age 20: What is known and what is not known about implicit bias. In *Behavior Research Methods*. <https://doi.org/10.31234/osf.io/bf97c>
- Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, 71(1), 419–445. <https://doi.org/10.1146/annurev-psych-010419-050837>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17–41. <https://doi.org/10.1037/a0015575>
- Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of Personality and Social Psychology*, 108(2), 187–218. <https://doi.org/10.1037/a0038557>
- Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., Tomezsko, D., Greenwald, A. G., & Banaji, M. R. (2019). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist*, 74, 569–586. <https://doi.org/10.1037/amp0000364>
- Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., Calanchini, J., Xiao, Y. J., Pedram, C., Marshburn, C. K., Simon, S., Blanchar, J. C., Joy-Gaba, J. A., Conway, J., Redford, L., Klein, R. A., Roussos, G., Schellhaas, F. M. H., Burns, M., ... Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology. General*, 145(8), 1001–1016. <https://doi.org/10.1037/xge0000179>
- Lai, C. K., & Wilson, M. E. (2021). Measuring implicit intergroup biases. *Social and Personality Psychology Compass*, 15(1), e12573. <https://doi.org/10.1111/spc3.12573>
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255–275. <https://doi.org/10.1037/0033-2909.125.2.255>
- Ma, D. S., Correll, J., Wittenbrink, B., Bar-Anan, Y., Sriram, N., & Nosek, B. A. (2013). When fatigue turns deadly: The association between fatigue and racial bias in the decision to shoot. *Basic and Applied Social Psychology*, 35(6), 515–524. <https://doi.org/10.1080/01973533.2013.840630>

- McFarland, S. (2010). Authoritarianism, social dominance, and other roots of generalized prejudice. *Political Psychology*, 31(3), 453–477. <https://doi.org/10.1111/j.1467-9221.2010.00765.x>
- Murphy, A., & Hine, B. (2019). Investigating the demographic and attitudinal predictors of rape myth acceptance in U.K. Police officers: Developing an evidence-base for training and professional development. *Psychology, Crime & Law*, 25(1), 69–89. <https://doi.org/10.1080/1068316X.2018.1503663>
- 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. <https://doi.org/10.1037/a0032734>
- Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual Review of Psychology*, 72(1), 533–560. <https://doi.org/10.1146/annurev-psych-071620-030619>
- Payne, B. K., & Correll, J. (2020). Race, weapons, and the perception of threat. *Advances in Experimental Social Psychology*, 62, 1–50.
- Peak, S. A., Hanson, E. J., Eadeh, F. R., & Lambert, A. J. (2016). On the light versus dark side of empathy: Implications for intergroup dynamics in a diverse society. In *The crisis of race in higher education: A day of discovery and dialogue*. Emerald Group Publishing Limited.
- Penner, L. A., Blair, I. V., Albrecht, T. L., & Dovidio, J. F. (2014). Reducing racial health care disparities: A social psychological analysis. *Policy Insights from the Behavioral and Brain Sciences*, 1(1), 204–212. <https://doi.org/10.1177/2372732214548430>
- Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, 49(1), 65–85. <https://doi.org/10.1146/annurev.psych.49.1.65>
- Plant, E. A., Goplen, J., & Kunstman, J. W. (2011). Selective responses to threat: The roles of race and gender in decisions to shoot. *Personality and Social Psychology Bulletin*, 37(9), 1274–1281. <https://doi.org/10.1177/0146167211408617>
- Pratto, F., Sidanius, J., & Levin, S. (2006). Social dominance theory and the dynamics of intergroup relations: Taking stock and looking forward. *European Review of Social Psychology*, 17(1), 271–320.
- Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences*, 114(41), 10870–10875. <https://doi.org/10.1073/pnas.1706255114>
- Racial and Identity Profiling Advisory Board. (2021). *Racial and Identity Profiling Advisory Board 2021 annual report*.
- Rooth, D.-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, 17(3), 523–534. <https://doi.org/10.1016/j.labeco.2009.04.005>
- Sadler, M. S., Correll, J., Park, B., & Judd, C. M. (2012). The world is not black and white: Racial bias in the decision to shoot in a multiethnic context. *Journal of Social Issues*, 286–313. [https://doi.org/10.1111/j.1540-4560.2012.01749.x@10.1111/\(ISSN\)1540-4560.POLICINGANDRACE](https://doi.org/10.1111/j.1540-4560.2012.01749.x@10.1111/(ISSN)1540-4560.POLICINGANDRACE)

- Schuman, H., Steeh, C., Bobo, L. D., & Krysan, M. (1997). *Racial attitudes in America: Trends and interpretations*. Harvard University Press.
- Shih, M., Wang, E., Trahan Bucher, A., & Stotzer, R. (2009). Perspective taking: Reducing prejudice towards general outgroups and specific individuals. *Group Processes & Intergroup Relations*, 12(5), 565–577. <https://doi.org/10.1177/1368430209337463>
- Sidanius, J., Cotterill, S., Sheehy-Skeffington, J., Kteily, N., & Carvacho, H. (2017). Social dominance theory: Explorations in the psychology of oppression. In *The Cambridge handbook of the psychology of prejudice* (pp. 149–187). Cambridge University Press. <https://doi.org/10.1017/9781316161579.008>
- Sidanius, J., Liu, J. H., Shaw, J. S., & Pratto, F. (1994). Social dominance orientation, hierarchy attenuators and hierarchy enhancers: Social Dominance Theory and the criminal justice system. *Journal of Applied Social Psychology*, 24(4), 338–366. <https://doi.org/10.1111/j.1559-1816.1994.tb00586.x>
- Sim, J. J., Correll, J., & Sadler, M. S. (2013). Understanding police and expert performance: When training attenuates (vs. exacerbates) stereotypic bias in the decision to shoot. *Personality & Social Psychology Bulletin*, 39(3), 291–304. <https://doi.org/10.1177/0146167212473157>
- Son Hing, L. S., Chung-Yan, G. A., Hamilton, L. K., & Zanna, M. P. (2008). A two-dimensional model that employs explicit and implicit attitudes to characterize prejudice. *Journal of Personality and Social Psychology*, 94(6), 971–987. <https://doi.org/10.1037/0022-3514.94.6.971>
- Stephan, W. G., & Finlay, K. (1999). The role of empathy in improving intergroup relations. *Journal of Social Issues*, 55(4), 729–743. <https://doi.org/10.1111/0022-4537.00144>
- Swencionis, J. K., & Goff, P. A. (2017). The psychological science of racial bias and policing. *Psychology, Public Policy, and Law*, 23(4), 398–409. <https://doi.org/10.1037/law0000130>
- Swencionis, J. K., Pouget, E. R., & Goff, P. A. (2021). Supporting social hierarchy is associated with White police officers' use of force. *Proceedings of the National Academy of Sciences*, 118(18). <https://doi.org/10.1073/pnas.2007693118>
- Voigt, R., Camp, N. P., Prabhakaran, V., Hamilton, W. L., Hetey, R. C., Griffiths, C. M., Jurgens, D., Jurafsky, D., & Eberhardt, J. L. (2017). Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences*, 114(25), 6521–6526. <https://doi.org/10.1073/pnas.1702413114>
- 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*.
- Xu, K., Nosek, B. A., & Greenwald, A. (2014). Psychology data from the race implicit association test on the project implicit demo website. *Journal of Open Psychology Data*, 2(1), e3. <https://doi.org/10.5334/jopd.ac>