Reducing Implicit Prejudice

Calvin K. Lai
Kelly M. Hoffman
Brian A. Nosek

University of Virginia

Authors’ note: Address correspondence to clai@virginia.edu; Box 400400, 102 Gilmer Hall, Charlottesville, VA 22904-4400. This project was supported by Project Implicit, Inc. Nosek is an officer of Project Implicit, a non-profit organization that includes in its mission “To develop and deliver methods for investigating and applying phenomena of implicit social cognition, including especially phenomena of implicit bias based on age, race, gender or other factors.”

Abstract

Reducing implicit prejudice is of interest for both basic and applied reasons. Basic research examines the mechanisms of altering implicit evaluation such as changing the associations, altering the surrounding context, and controlling the activation or expression of associations. Applied research examines the effectiveness of interventions to reduce implicit prejudice and, ultimately, discrimination. There is substantial evidence identifying multiple mechanisms for reducing implicit prejudice, but much less evidence of the effectiveness of these interventions in applied contexts. Further, there is substantial evidence for the short term malleability of implicit prejudice, but much less evidence for long-term change. Finally, there is substantial evidence of the link between implicit prejudice and behavior, but no evidence yet for whether changing implicit prejudice leads to changes in behavior. These knowledge gaps provide a clear agenda for research on reducing implicit prejudice.

Abstract = 136 words
Reducing Implicit Prejudice

Gordon Allport memorably defined prejudice as a “feeling, favorable or unfavorable, toward a person or thing, prior to, or not based on, actual experience” (Allport, 1954, p. 6). The conceptualization of prejudice has evolved since with a focus on intergroup relations – evaluations of others based on race, ethnicity, gender, social class, sexual orientation, nationality, religion, or disability (Brewer, 1999; Dovidio, Glick & Rudman, 2005; Tajfel, 1982; Yzerbyt & Demoulin, 2010). One important shift in the understanding of prejudice was the recognition that the “feeling” need not be deliberate, intentional, endorsed, or even available to conscious awareness (Fazio, Jackson, Dunton, & Williams, 1995; Greenwald & Banaji, 1995; Banaji, Nosek, & Greenwald, 2004). People can have implicit prejudices – feelings, favorable or unfavorable, toward persons or groups that they did not intend, endorse, or even realize that they possessed (Gawronski & Payne, 2010; Nosek, Hawkins, & Frazier, 2011, 2012).

Implicit prejudice is distinct, but related to explicit prejudice (Nosek & Smyth, 2007). For example, self-reported attitudes toward Blacks compared to Whites are moderately positively correlated with an Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998; Nosek, Greenwald, & Banaji, 2007) measuring associations for the same racial groups (zero-order correlation r ~ .30; latent variable correlation r ~ .45; see Nosek, 2007 for a review). But, the strength of the relationship between implicit and explicit attitudes varies across social categories with age and disability attitudes eliciting particularly weak relationships (r’s < .15), for example, and sexual orientation and political attitudes eliciting comparatively strong ones (r’s > .45; Nosek, Smyth, et al., 2007). This variation may be explained by, among other factors, the social pressures against holding negative attitudes toward some groups, and how much people have elaborated on those attitudes (Nosek, 2005). Further, both implicit and explicit evaluations predict behavior – jointly and independently (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Finally, implicit and explicit evaluations are understood to be subject to distinct formative experiences (Ranganath & Nosek, 2008; Ratliff & Nosek, 2011; Rydell & McConnell, 2006), operate via distinct psychological mechanisms (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009), and, as a consequence, have distinct routes for change (Gawronski & Bodenhausen, 2006).

Basic Research on Reducing Implicit Prejudice: Mechanisms of Change

There are at least three means of altering the expression of implicit prejudice: retraining the underlying associations, shifting the context of evaluation, and controlling the activation or
application (Gawronski & Sritharan, 2010). Here we provide a summary review organized by these categories (see also Dasgupta, 2009; Sritharan & Gawronski, 2010).

Retraining Associations.

Evaluative conditioning. Implicit attitudes and stereotypes, of which implicit prejudice is a special case, are understood to reflect associations between concepts (e.g., male/female, Black/White, old/young) and evaluations (e.g., good/bad, smart/dumb; Greenwald et al., 2002). Perhaps the most direct method to change these associations is evaluative conditioning (Bar-Anan, De Houwer, & Nosek, 2010; De Houwer, Thomas, & Baeyens, 2001; Karpinski & Hilton, 2001; Olson & Fazio, 2001, 2006). Evaluative conditioning provides experience linking concepts with associations that differ from their preexisting association to retrain or create alternative associations. For example, Olson and Fazio (2006) briefly presented participants with 24 positive images and words paired with Black faces, and 24 negative images and words paired with White faces. Exposure to these pairings reduced implicit racial prejudice immediately, and this change persisted at a follow-up assessment two days later.

The logic of retraining associations appears in other approaches. For example, on the assumption that people tend to approach things that are good and avoid things that are bad, a computer exercise with 480 trials of initiating approach toward Black faces and avoidance of White faces reduced implicit prejudice (Kawakami, Phillips, Steele, & Dovidio, 2007; see also Wennekers, Holland, Wigboldus, & Knippenberg, 2011). This change may have been due to the self – a concept strongly associated with good (Greenwald et al., 2002) – becoming more associated with the approached faces (Kawakami, Steele, Cifa, Phillips, & Dovidio, 2008; Phillips, Kawakami, Tabi, Nadolny, & Inzlicht, 2011). Another approach is to practice responding “Yes” to pairings of Black people with counterstereotypical words (e.g., “intelligent”; Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000). The complementary approach of saying “No” to pairings of Black targets with stereotypical words does not appear to be effective (Gawronski, Deutsch, Mbirkou, Seibt, & Strack, 2008). Negations require first representing the association and then invalidating it. Invalidating an association requires additional processing (Gilbert, Taforadi, & Malone, 1993) and may be particularly difficult to process implicitly (Gawronski et al., 2008).

Intergroup contact. Intergroup contact is the most well studied means of reducing explicit prejudice (Allport, 1954; Pettigrew & Tropp, 2006), and it also appears to reduce implicit prejudice (Aberson, Porter, & Gaffney, 2008; Aberson, Shoemaker, & Tomolillo, 2004; Dasgupta & Rivera, 2008). Whereas explicit prejudice reduction is better predicted by the quality of contact, implicit prejudice reduction is better predicted by the quantity of contact (Tam, Hewstone, Harwood, Voci, & Kenworthy, 2006; Turner, Hewstone, & Voci, 2007). Additionally, recent research suggests that just imagining positive intergroup contact can reduce implicit prejudice (Turner & Crisp, 2010; Vezzali, Capoza, Giovannini, & Stathi, 2011). Most contact research uses correlational or quasi-experimental designs. A notable exception took

---

1 These do not commit to a particular mental model. From a representational perspective (Petty & Brinol, 2006; Petty, Brinol, & DeMarree, 2007; Wilson, Lindsey, & Schooler, 2000), these can be understood as changing the association itself, shifting to related associations, or altering the expression of associations. From a distributed network or connectionist perspective (Smith, 1996, 2009; Conrey & Smith, 2007, Mitchell, Nosek, & Banaji, 2003; Monroe & Read, 2008), these can be understood as changing the connection weights in the network, altering the activated nodes, and altering the network output. Both perspectives can be adapted to accommodate virtually any findings.
advantage of a natural experiment of college roommate assignments (Shook & Fazio, 2008). Whites randomly assigned to live with a Black roommate had lower implicit prejudice after one semester than students who were randomly assigned to live with a White roommate.

**Persuasion.** Persuasive appeals have been extensively studied for changing explicit attitudes (Eagly & Chaiken, 1993; Petty & Wegener, 1998), but not implicit attitudes. One exception showed that an argument eliciting high cognitive elaboration for integrating more African-American professors into a university led to weaker implicit racial prejudice compared to an argument eliciting low cognitive elaboration (Brinol, Petty, & McCaslin, 2009). Cognitive elaboration may influence implicit attitudes through propositional reasoning, whereby newly-gained knowledge leads to the creation of positive associations with attitude objects. Further, Marini, Rubichi, and Sartori (in press) found that the personal relevance of a counterstereotypical message affected the effectiveness of the message on implicit prejudice. There is some evidence that other persuasion cues such as source expertise and trustworthiness can change implicit associations (Smith, De Houwer, & Nosek, 2012), but this has not been tested with implicit prejudice.

**Shifting the Context.**

**Counterstereotypical exemplars.** The context of evaluation plays an important role in the expression of implicit prejudice. Implicit evaluations are sensitive to shifts in the representation of the target group, the mindset of evaluation, the interpersonal context, and affective states.

One approach is to elaborate on the counterstereotypical associations of positive exemplars of disliked groups and negative exemplars of liked groups (Dasgupta & Asgari, 2004; Dasgupta & Greenwald, 2001; Dasgupta & Rivera, 2008; Gonsalkorale, Allen, Sherman, & Klauer, 2010; Joy-Gaba & Nosek, 2010). For example, Dasgupta and Greenwald (2001) demonstrated that exposing participants to admired African American exemplars and disliked European American exemplars and making their ethnicity salient reduced implicit preferences for Whites compared to Blacks. Similarly, engaging in counterstereotypic mental imagery (Blair, Ma, & Lenton, 2001; Hugenberg, Blusierwicz, & Sacco, 2010), or presenting negative information about the ingroup as a contrast to the disliked outgroup (Sassenberg & Wieber, 2005) can reduce implicit prejudice.

A similar approach shifts the representation of the target group not through elaboration, but by changing the perceived target of evaluation. For example, changing the stimuli to include counterstereotypical exemplars (Govan & Williams, 2004) or positive stereotype-consistent features of an outgroup (Rodriguez-Bailon, Ruiz, & Moya, 2009) elicit less implicit prejudice. Likewise, Black targets are evaluated more positively and less stereotypically when evaluated by occupation (Mitchell, et al., 2003) or by age (Jones & Fazio, 2010) rather than by race. Presenting targets within particular social roles and environments also influences implicit evaluations. Showing a prison context with a Black target dressed as a lawyer elicited more positive evaluations compared to when that same Black target was in the role of prisoner (Barden, Maddux, Petty, and Brewer, 2004; see also Maddux, Barden, Brewer, & Petty, 2005). Similarly, Black targets are evaluated more positively when the targets are placed in front positive backgrounds (e.g., a family barbeque) compared to negative backgrounds (a gang incident; Wittenbrink, Judd, & Park, 2001).

**Ideologies, goals, and motivations.** Ideologies, goals, and motivations can influence how people implicitly evaluate outgroup members. Priming participants with a multicultural
ideology elicited less implicit racial prejudice compared with a colorblind ideology (Richeson & Nussbaum, 2004; but see Lai, Smyth, & Nosek, 2012, and Correll, Park, Allegra, & Smith, 2008 for limiting conditions). Also, blurring intergroup boundaries by asking participants to generate characteristics they have in common with an outgroup reduces implicit prejudice (Hall, Crisp, & Suen, 2009). Other studies suggest the effectiveness of priming particular goals or motivations. Participants informed that they were not progressing on goals to be egalitarian had lower levels of implicit racial prejudice relative to participants who were told they were becoming more egalitarian, perhaps because the latter felt that they no longer needed to pursue the goal (Mann & Kawakami, 2012; see also Legault, Gutsell, & Inzlicht, 2011). Similarly, priming egalitarian goals led to greater stereotype inhibition than a control goal, however, this did not occur for participants whose egalitarian goals had been affirmed (Moskowitz & Li, 2011). For an egalitarian goal to be effective, it may be a necessary to activate it while avoiding any available evidence that it has been achieved.

People automatically adapt their self-concepts, beliefs, and attitudes to the demands of the interpersonal context (Lowery, Hardin, & Sinclair, 2001; Lun, Sinclair, Whitchurch, & Glenn, 2007; Sinclair, Lowery, Hardin, & Colangelo, 2005). As a result, people may “socially tune” to the perceived attitudes of others in intergroup situations. In one study, participants who first interacted with a Black experimenter later exhibited lower implicit prejudice than participants who interacted with a White experimenter, suggesting that automatic responses shifted to facilitate smooth interaction with the Black experimenter (Lowery et al., 2001). Furthermore, the mere physical presence of others in the same experiment room can decrease implicit prejudice and increased the accessibility of egalitarian-related concepts (Castelli & Tomelleri, 2008). Finally, social roles invoking mindsets about power dynamics influences the expression of implicit prejudice. In one study, Whites who anticipated interacting with a Black superior showed lower levels of implicit prejudice compared to Whites who anticipated interacting with a Black subordinate (Richeson & Ambady, 2003; see also Richeson & Ambady, 2001).

**Affect.** Attitudes contain an affective component, and are subject to influence by emotional states. Consistent with evidence that accessible associations are promoted by positive moods and inhibited by negative moods (Clore & Huntsinger, 2007; Clore et al., 2001), people in negative moods show less implicit prejudice than people in positive moods (Huntsinger, Sinclair, & Clore, 2009). However, positive mood can lead to less implicit stereotyping for individuals with chronically accessible egalitarian goals (Huntsinger, Sinclair, Dunn, & Clore, 2010). Beyond mood, specific emotions can influence implicit prejudice. For example, anger and disgust can increase implicit prejudice for groups associated with those emotions (i.e., anger for Arabs, and disgust for gay people; Dasgupta, DeSteno, Williams, & Hunsinger, 2009) and for unfamiliar groups (Dasgupta et al., 2009; DeSteno, Dasgupta, Bartlett, & Cajdric, 2004), but no published research demonstrates an emotion that reduces implicit prejudice. However, administering Propranolol, a medication used to lower blood pressure and anxiety, also elicited less implicit prejudice compared to a placebo group, perhaps because it reduces affective responses to threat (Terbeck et al., in press).

**Behavioral strategies altering activation or application.** A person may possess strong implicit associations linking gay people with bad and straight people with good, but nonetheless not show a strong effect on an implicit measure, or on a behavioral measure because they somehow prevent activation or application of those associations (Conrey, Sherman, Gawronski,
Hugenberg, & Groom, 2005). Shifts due to behavioral strategies may not affect the existence of implicit prejudice, but they could alter its expression.

Manipulations that impair control over automatic responses can increase implicit prejudice and stereotyping include alcohol consumption (Bartholow, Dickter, & Sestir, 2006), ego-depletion (Govorun & Payne, 2006), and anticipation that responses on an implicit measure will be shared with others (Lambert et al., 2003). Other manipulations may enhance control or provide effective strategies for preventing expression of implicit prejudice, such as giving instructions to avoid stereotyping or prejudice (Lowery et al., 2001; Wallaert, Ward, & Mann, 2010) or providing implementation intentions (Mendoza, Gollwitzer, & Amodio, 2010; Stewart & Payne, 2008; Webb, Sheeran, & Pepper, 2012). Implementation intentions are “if-then” plans that link a situational cue to a behavioral response (Gollwitzer, 1999), increasing the automaticity of behavior by increasing the consistency between goal-directed intentions and behavior. Establishing a plan to think “good” after seeing a Black face decreased implicit racial prejudice compared to a plan to think “quickly” after seeing a Black face (Stewart & Payne, 2008). Further, chronic motivations to avoid prejudiced responding may enhance the ability to override automatic reactions (Allen, Sherman, & Klauer, 2010; Legault, et al., 2009; Maddux, et al., 2005; Moskowitz, et al., 1999).

Finally, participants may adopt behavioral strategies that interfere with implicit measurement itself. For example, participants may misunderstand or deliberately ignore task instructions and thus appear to have different implicit prejudices than people who understand and follow task instructions. For example, participants directly evaluating the target concepts instead of categorizing them in the personalized Implicit Association Test (Nosek & Hansen, 2008) and Affect Misattribution Procedure (Bar-Anan & Nosek, in press) compromises the interpretation of the method as an implicit measure. Further, while participants do not appear to spontaneously fake the Implicit Association Test or the evaluative priming task easily (Banse, Seise, & Zerbes, 2001; Kim, 2003), providing experience and specific instructions is effective for facilitating faking behavior (Czellar, 2006; Degner, 2009; De Houwer, Beckers, & Moors, 2007; Fiedler & Bluemke, 2005; Klauer & Teige-Mocigemba, 2007; Steffens, 2004; Teige-Mocigemba & Klauer, 2008; Verschuere, Prati, & De Houwer, 2009). Some faking is statistically detectable and correctable (Cvencek, Greenwald, Brown, Gray, & Snowden, 2010). Even so, attention to measurement of irrelevant behavioral strategies, like faking, is important for accurately inferring the alteration of implicit evaluations.

Evidence for change without clarity for what is changing. The prior sections seem to imply that changing associations, contextual shifts, and behavioral strategies preventing activation or application are clearly distinct, and that each demonstration fits into one and only one domain. That is not correct. In many cases, the categorization of particular manipulations is done via an estimation of plausibility, not direct empirical evidence. Also, any given manipulation could leverage multiple mechanisms at one time. Part of investigating the psychological mechanisms of reducing implicit prejudice is to translate the operational features of manipulations into the theoretical psychological processes that they influence.

The majority of mechanisms for malleability in the research literature are qualitatively determined. For instance, taking the perspective of outgroup members decreases implicit prejudice and stereotyping (Galinsky & Moskowitz, 2000; Todd, Bodenhausen, Richeson, & Galinsky, 2011), but it is unclear as to whether it temporarily modifies activation of associations or if it changes associations, or even if it induces behavioral control strategies to prevent expression of implicit prejudice. Likewise, associations can form and change very rapidly by
assimilating them with another well-elaborated concept, like political attitudes (Smith, Ratliff, & Nosek, 2012), but it is not clear whether this is better understood as retraining or change in context. Experimental ingenuity, analytic techniques like QUAD modeling (Conrey et al., 2005), and improvements in the taxonomy of psychological processes will help further clarify the mechanisms of implicit prejudice reduction.

**Applied Research on Reducing Implicit Prejudice: What Actually Works?**

Can implicit evaluations be changed? The evidence from the preceding section demonstrates that the answer is yes. However, saying whether something can be true is a small claim. From a perspectivist approach to science, all claims are true – if only in very limited circumstances (McGuire, 1973). For applied interests, the question is not “Can an effect occur?” The question is “Does the effect occur?” Is the intervention actually effective, to what degree, and under what conditions? Identifying mechanisms of change does not, on its own, provide insight primarily because the emphasis is on isolating mechanisms for theoretical purposes, not examining their effectiveness for comparative or practical purposes.

Compared to the identification of mechanisms, much less is known about the effectiveness and boundary conditions of interventions to reduce implicit prejudice. In this section, we summarize the evidence by asking five questions that do not yet have clear answers. While the questions have basic and applied implications, we focus on their implications for reducing implicit prejudice.

**Malleability or change?** Malleability refers to shifts in evaluation that are limited to the immediate situational context; change refers to shifts in evaluation that persist across multiple situational contexts (Nosek et al., 2012). For many applied interests, interventions must have lasting consequences beyond the immediate circumstance of the intervention. The three categories in the prior section imply whether they demonstrate malleability versus change: altering associations would presumably have general, long-term consequences, while contextual changes and behavioral strategies altering expression would have specific, short-term consequences. But, the evidence for malleability versus change of particular interventions is lacking. For example, of the 174 studies from articles cited in this paper that sought evidence for change in implicit associations, just 12 (6.9%) assessed implicit evaluations in a time other than the original intervention session. Most studies assessed change immediately following the manipulation in the same experimental context. This prevents any inference about whether the effects reflect situational, short-term malleability or general, long-term change.

There is some evidence for long-term change in implicit prejudice. In the laboratory, researchers has demonstrated that exposure to counterstereotypical exemplars (Dasgupta & Greenwald, 2001), evaluative conditioning (Olson & Fazio, 2006), and approach-avoidance training (Kawakami et al., 2000) reduce implicit prejudice immediately and for at least one to two days later in the same context. Outside the laboratory, a seminar on prejudice and intergroup conflict was associated with reduced implicit prejudice at the end of a semester (Rudman, Ashmore, & Gary, 2001), implementation intentions was associated with reduced implicit gender-leader stereotypes after 3 weeks (Webb et al., 2012), and having an outgroup roommate was related to decreased implicit prejudice after one school quarter (Shook & Fazio, 2008). Dasgupta and Asgari (2004) extended laboratory findings on counterstereotypical exemplars in a natural experiment, showing that undergraduate women who had more contact with female instructors during their first year at college held weaker implicit stereotypes associating leadership with men than women who had less contact with female instructors, one year after the
exposure. However, Stout, Dasgupta, Hunsinger, and McManus (2011) found that being assigned to female math professors and teaching assistants led to more positive implicit attitudes toward math and greater identification with math, but not reduced implicit stereotypes associating math with men over the course of a semester. As far as we are aware, these studies comprise every demonstration of change in implicit prejudice that holds across time (see Joy-Gaba & Nosek, 2012, O’Brien et al., 2010, Marini et al., in press, for examples of lack of change over time), and for four of them, the intervention was ongoing at assessment making it unclear whether the shift was lasting. In sum, the existing literature provides solid evidence for implicit prejudice malleability, but little and mixed evidence for implicit prejudice change. With the relatively compelling theoretical rationale and evidence from other domains, we speculate that the paucity of evidence is a function of the feasibility challenges for conducting longitudinal research rather than the unchangeability of implicit prejudice.

Does publishing bias exaggerate evidence for malleability? Publishing practices may limit the identification of constraints and boundary conditions for interventions to reduce implicit prejudice. In particular, achieving a positive result – i.e., demonstrating that a manipulation changes implicit evaluation – is a de facto requirement for publication (Fanelli, 2010, 2012; Greenwald, 1975; Nosek, Spies, & Motyl, in press; Sterling, 1959; Sterling, Rosenbaum, & Weinkam, 1995). The consequence is that the published literature contains much less evidence of the conditions that do not elicit change in implicit prejudice, even though they are sure to exist – if only in the file drawer (Rosenthal, 1979). Of the 174 studies from articles cited in this paper that sought evidence for change in implicit associations, just 18 (10.3%) reported evidence of no change on implicit evaluations. Notably, 37 studies (21.3%) showed a moderator effect in which change was observed in one condition but not in another condition. In one sense that suggests effective condition seeking; in another sense, particularly with so few examples of null main effects, it suggests that some apparent moderators may have been identified in exploratory analysis. Moderators identified in this way are more likely to be false positives (Simmons, Nelson, & Simonsohn, 2011).

A common rationalization for the positivity bias in publishing is that there are many reasons for a study to go “wrong” – i.e., show no difference where one exists – other than the intervention being ineffective. Of course, rote application of this rationalization prevents learning about conditions that are ineffective and guarantees that the published literature is an inaccurate reflection of reality. Further, there are also many reasons for a study to falsely go “right” – i.e., show a difference where there is not one (Simmons, et al., 2011; John, Loewenstein, & Prelec, 2012).

There is some evidence that the published literature overstates the evidence for malleability and change of implicit prejudice. For example, Plant and colleagues (2009) reported correlational evidence from two undergraduate samples (N’s = 219, 79) that Barack Obama’s presidential campaign reduced implicit racial prejudice to near neutrality. However, a replication by Schmidt and Nosek (2010) with a heterogeneous sample of 479,405 demonstrated a high degree of stability in implicit racial attitudes from four months prior to Obama’s announcement of his candidacy for president to four months after his inauguration. Further, a report of 39 attempts to reduce implicit preferences for Whites compared to Blacks found that 51.3% were effective (20 of 39; Lai et al., 2012). This was despite an average sample size of 251 (~80% to detect effects of $d=.26$ or larger), intervention designs based on published theory or existing evidence, and incentives for designers to have the most effective intervention (winning a contest). These null results are not easily dismissed for “uninteresting” reasons – e.g., sampling,
procedural and analytic details, implementation of the implicit measure – because other interventions were effective in identical conditions. The study design isolates the lack of effect to the interventions themselves (although it does not eliminate the possibility that the intervention was poorly implemented).

These results suggest that the pervasive tendency to underreport negative results applies to investigations of reducing implicit prejudice as well. It can be just as important to understand the circumstances in which an intervention is ineffective as when it is effective. For instance, Joy-Gaba and Nosek (2010) found limiting conditions for the effects of counterstereotypical exemplars. Adapting Dasgupta and Greenwald’s counterstereotype exposure paradigm (2001), they found that exposure to admired Black exemplars in isolation was not effective at reducing implicit racial prejudice. Other factors, such as simultaneous exposure to negative White exemplars or making race salient during the intervention, may be critical for inducing malleability in the paradigm. Likewise, for the earlier discussed social tuning effects, decreases in implicit prejudice were found only for participants who were motivated to affiliate or who interacted with a likable experimenter (Sinclair et al., 2005), who were epistemically-motivated (Lun et al., 2007), or who were in a positive mood (Huntsinger & Sinclair, 2010).

**What is actually effective?** The limitations of null hypothesis significance testing (NHST) have been well documented (Bakan, 1966; Cohen, 1994; Frick, 1996; Oakes, 1986), but it remains dominant practice for basic research perhaps because of its singular strength – establishing “sufficient evidence has been presented to support a claim, with sufficient defined as \( p < .05 \)” (Frick, 1996, p. 380). In prejudice reduction research, NHST answers questions about whether or not an intervention elicits less implicit prejudice than a control condition. However, NHST is not sufficient “for establishing beliefs or estimating the probability that these ordinal claims are correct.” (Frick, 1996, p. 380), nor does it estimate the magnitude of effects (Cohen, 1969). These latter concerns are vitally important for transforming the claims into effective interventions. In particular, knowing that an effect is statistically significant does not indicate that is practically significant (Cohen, 1994). For this, estimates of effect magnitude are critical.

Further, effect magnitude estimates for interventions in highly-controlled experimental contexts may bear little relationship to their effectiveness *in vivo*. Isolating causal variables in the “real world” can be difficult and unforgiving. The moderating effects of sample, setting, and eliciting conditions can eliminate or enhance the impact of manipulations exported from the laboratory (Willingham, 2012). Thus, for applied goals of actually reducing implicit prejudice, failing to conduct the experimental research in the applied context greatly limits the generalizability of phenomena found in the laboratory. While most of the research on changing implicit prejudice is based in the laboratory, a few examples demonstrate the potential for exportation and estimation of intervention effectiveness (Dasgupta & Asgari, 2004; O’Brien et al., 2010; Rudman et al., 2001; Shook & Fazio, 2008; Stout et al., 2011). A notable example took advantage of a natural experiment in Indian politics (Beaman, Chattopadhyay, Duflo, Pande, & Topalova, 2008). Village council leadership positions were randomly selected (by law) to be “reserved for women.” Men in villages that were required to have female council leaders held weaker implicit gender-leader stereotypes compared to men living in villages where no gender quota was established (but no difference was observed for implicit preferences for female leaders compared to male leaders).

**Does shifting implicit prejudice change behavior?** For applied research on reducing implicit prejudice, the actual target is not really implicit prejudice, but its presumed consequence—discrimination. If reducing implicit prejudice did not reduce discriminatory
Reducing Implicit Prejudice

behavior, then it is likely that applied interest in reducing implicit prejudice would fade. Policies and practices target behavior, not thoughts (Nosek & Riskind, 2012). This begs the question—does changing implicit prejudice change discriminatory behavior? Remarkably, there is very little direct empirical evidence to answer this question.

There is substantial evidence that implicit prejudice predicts discriminatory behavior (Greenwald et al., 2009; Jost et al., 2009). For instance, Rooth (2010) sent resumes with Swedish-sounding names or Arab-Muslim-sounding names to recruiters. Many of these recruiters also completed an IAT measuring implicit prejudice toward Arab-Muslims. Resumes with Arab-Muslim-sounding names were 45% less likely to be called back for an interview, and the likelihood of this hiring bias was predicted by implicit prejudice toward Arab-Muslims. This increases the plausibility that reducing implicit prejudice would therefore reduce discrimination. However, implicit prejudice was not experimentally manipulated.

There are a few studies that showed an intervention was effective at changing both implicit prejudice and behavior (Kawakami et al., 2007, 2008; Mann & Kawakami, 2012), but none reported whether implicit prejudice statistically mediated the effect of the manipulation on the behavior or, if they did report the analysis, there was no mediation (Mann & Kawakami, 2012).

We found no published study (whether successful or not) that tested whether a change in implicit prejudice predicted a later change in behavior. There is one published example of this critical step outside of prejudice research. Teachman and colleagues tracked changes in implicit panic associations and panic symptoms for individuals with a panic disorder during a 12-week cognitive behavior therapy program (Teachman, Marker, & Smith-Janik, 2008). They found significant reductions in both panic associations and behavioral symptoms over the course of treatment, and importantly, found that the earlier change in implicit panic associations predicted later change in panic symptoms. This provides evidence that changes in implicit associations can lead to behavior change, and is a model for testing whether this occurs for prejudice and discrimination.

Is it necessary to change implicit prejudice? The preceding section suggests an even more fundamental question—is it necessary to reduce implicit prejudice to reduce discrimination? Surely the answer is no, at least under some conditions. For example, gender discrimination in orchestral hiring is reduced by instituting blind auditions (Goldin & Rouse, 1997). Blind evaluation does not change the perceiver’s thoughts or feelings about gender; it eliminates the possibility of using gender as a basis of evaluation. As such, interventions need not change implicit or explicit biases—they can just address their expression.

Extending the orchestral example, eliminating information about age, race, or sexual orientation from the decision-making context makes it difficult to use any social biases about those categories in social judgment. Providing objective criteria to guide decision-making, such as listing job requirements immediately prior to selecting a candidate (Uhlmann & Cohen, 2005), can constrain subjective decision-making. And, providing insight that one’s behavior could be influenced by implicit biases can instigate efforts to control the expression of such biases without affecting the biases themselves (Bartlett, 2009; Monteith, Ashburn-Nardo, Voils, & Czopp, 2002; Monteith & Mark, 2005; but see Joy-Gaba & Nosek, 2012 for evidence that knowledge alone may not be sufficient). However, mindset interventions can also increase discrimination, sometimes ironically. For example, inducing people to assert that they are objective decision-makers prior to a hiring decision increases gender (Uhlmann & Cohen, 2007) and age discrimination (Lindner, Nosek, & Graser, 2012). This emphasizes the delicacy of inducing
mindsets to alter behavior — whereas the motivation to be non-prejudiced may lead to reduced discrimination (Bartlett, 2009; Plant, Devine, & Peruche, 2010), thinking of oneself as non-prejudiced may ironically increase discrimination (Monin & Miller, 2001).

So, if the potential effects of implicit prejudice on behavior can be addressed without changing it, why investigate ways to reduce implicit prejudice? For one, it is possible that reducing implicit prejudice is more effective than many of the other strategies, and — for some circumstances — might be the only available strategy. For example, in some performance evaluation contexts it may not be feasible to blind decision makers, eliminate subjectivity, provide sufficient accountability, or induce a strong enough evaluative mindset to eliminate the potential for implicit biases to be expressed. Also, it is conceivable that interventions on the decision-making process itself are effective as short-term, local interventions, but that implicit prejudice reduction is more effective as a long-term, general intervention strategy — particularly when there is little control of the daily decision-making processes. In any case, identifying the comparative effectiveness and boundary conditions for interventions that target the prejudice versus the expression of prejudice will facilitate clarification of the most effective means of having the desired impact on behavior.

Conclusion

Research on implicit prejudice illustrates the happy marriage between basic and applied approaches. Basic research examines mechanisms for changing implicit cognitions and provides insight into the operations of the mind that escape conscious awareness or control. Applied research investigates how to reduce implicit prejudice with a goal of mitigating discriminatory behavior. The discovery of effective mechanisms for change provides input for the development of interventions; and, evidence for effective interventions feeds back into investigations of the operative causes. In particular, effective interventions may leverage multiple mechanisms simultaneously providing much greater impact than they would individually. Basic investigation of the simultaneous and interactive effects of multiple mechanisms on behavior is daunting because of the practical challenges of exerting systematic control over many variables at once. Applied evidence can simplify the search by showing interventions that work, even if we do not yet know how.
References


Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90, 891-904.


