Journal of Experimental Psychology: General

0096-3445/14/$12.00 DOI: 10.1037/a0036260

Reducing Implicit Racial Preferences: I. A Comparative Investigation of 17 Interventions

Calvin K. Lai
University of Virginia

Steven A. Lehr and Carlo Cerruti
Harvard University

Jennifer A. Joy-Gaba
Virginia Commonwealth University

Bethany A. Teachman
University of Virginia

Spassena P. Koleva
University of Southern California

Larisa Heiphetz
Boston College

Rhiannon N. Turner
Queen's University Belfast

Selin Kesebir
London Business School

Sandro Rubichi
University of Modena and Reggio Emilia

Christopher M. Dial
Harvard University

Mahzarin R. Banaji
Harvard University

Maddalena Marini
University of Modena and Reggio Emilia

Jiyun-Elizabeth L. Shin
Stony Brook University

Arnold K. Ho
Colgate University

Sean P. Wojcik
University of California, Irvine

Rebecca S. Frazier
University of Virginia

Eva E. Chen
The Hong Kong University of Science and Technology

Jonathan Haidt
New York University

Carlee Beth Hawkins and Hillary S. Schaefer
University of Virginia

Giuseppe Sartori
University of Padua

N. Sriram
University of Virginia

Brian A. Nosek
University of Virginia

Calvin K. Lai, Department of Psychology, University of Virginia; Maddalena Marini, Department of Psychology, University of Modena and Reggio Emilia; Steven A. Lehr and Carlo Cerruti, Department of Psychology, Harvard University; Jiyun-Elizabeth L. Shin, Department of Psychology, Stony Brook University; Jennifer A. Joy-Gaba, Department of Psychology, Virginia Commonwealth University; Arnold K. Ho, Department of Psychology, Colgate University; Bethany A. Teachman, Department of Psychology, University of Virginia; Sean P. Wojcik, Department of Psychology, University of California, Irvine; Spassena P. Koleva, Department of Psychology, University of Southern California; Rebecca S. Frazier, Department of Psychology, University of Virginia; Larisa Heiphetz, Department of Psychology, Boston College; Eva E. Chen, Department of Psychology, The Hong Kong University of Science and Technology; Rhiannon N. Turner, Department of Psychology, Queen's University Belfast; Jonathan Haidt, Stern School of Business, New York University; Selin Kesebir, London Business School; Carlee Beth Hawkins and Hillary S. Schaefer, Department of Psychology, University of Virginia; Sandro Rubichi, Department of Psychology, University of Modena and Reggio Emilia; Giuseppe Sartori, Department of Psychology, University of Padua; Christopher M. Dial; Department of Psychology, Harvard University; N. Sriram, Department of Psychology, University of Virginia; Mahzarin R. Banaji, Department of Psychology, Harvard University; Brian A. Nosek, Department of Psychology, University of Virginia.

This project was supported by a gift from Project Implicit. Calvin K. Lai and Carlee Beth Hawkins are consultants, and Brian A. Nosek is an officer of Project Implicit, Inc., a nonprofit 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." Calvin K. Lai and Brian A. Nosek conceived the current research; all authors designed and performed the current research; Calvin K. Lai analyzed the data.

Correspondence concerning this article should be addressed to Calvin K. Lai, Department of Psychology, University of Virginia, Charlottesville, VA 22904. E-mail: ckl5ae@virginia.edu

Many methods for reducing implicit prejudice have been identified, but little is known about their relative effectiveness. We held a research contest to experimentally compare interventions for reducing the expression of implicit racial prejudice. Teams submitted 17 interventions that were tested an average of 3.70 times each in 4 studies (total $N = 17{,}021$), with rules for revising interventions between studies. Eight of 17 interventions were effective at reducing implicit preferences for Whites compared with Blacks, particularly ones that provided experience with counterstereotypical exemplars, used evaluative conditioning methods, and provided strategies to override biases. The other 9 interventions were ineffective, particularly ones that engaged participants with others' perspectives, asked participants to consider egalitarian values, or induced a positive emotion. The most potent interventions were ones that invoked high self-involvement or linked Black people with positivity and White people with negativity. No intervention consistently reduced explicit racial preferences. Furthermore, intervention effectiveness only weakly extended to implicit preferences for Asians and Hispanics.

Keywords: attitudes, racial prejudice, implicit social cognition, malleability, Implicit Association Test

Thoughts and feelings outside of conscious awareness shape social perception, judgment, and action (Bargh, 1999; Devine, 1989; Greenwald & Banaji, 1995). Nowhere has this idea been more explored than in studies of racial prejudice in which people report egalitarian racial attitudes, but also implicitly prefer Whites compared with Blacks (Devine, 1989; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Jackson, Dunton, & Williams, 1995; Nosek, Smyth, et al., 2007). These studies have been influential because implicit racial preferences predict behaviors such as negative interracial contact (McConnell & Leibold, 2001), biases in medical decision making (Green et al., 2007), and hiring discrimination (Rooth, 2010). From the hundreds of studies conducted, we can conclude that implicit preferences (a) are related to, but distinct from, explicit preferences (Greenwald & Banaji, 1995; Nosek & Smyth, 2007), (b) are constructed through different mechanisms than explicit preferences (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009; Ranganath & Nosek, 2008; Ratliff & Nosek, 2011; Rydell & McConnell, 2006), and (c) have distinct mechanisms for change compared with explicit preferences (Gawronski & Bodenhausen, 2006).

The study of implicit racial preferences has been driven not only by an interest in the basic mechanisms underlying social cognition but also by an applied interest in social change: How can the expression of implicit racial preferences be reduced to mitigate subsequent discriminatory behavior? Indeed, significant progress has been made in the goal of identifying the processes underlying malleability and change in implicit evaluations (Dasgupta & Greenwald, 2001; Mitchell, Nosek, & Banaji, 2003; Olson & Fazio, 2006; Rudman, Ashmore, & Gary, 2001; for reviews, see Blair, 2002; Dasgupta, 2009; Gawronski & Bodenhausen, 2006; Gawronski & Sritharan, 2010; Lai, Hoffman, & Nosek., 2013; Sritharan & Gawronski, 2010). From such research, we know the expression of implicit racial preferences can be shifted through changes in emotional states (Dasgupta, DeSteno, Williams, & Hunsinger, 2009; DeSteno, Dasgupta, Bartlett, & Cajdric, 2004), exposure to counterstereotypical exemplars (e.g., Dasgupta & Greenwald, 2001), and setting egalitarian goals (Legault, Gutsell, & Inzlicht, 2011; Mann & Kawakami, 2012).

This basic research on changes in the expression of implicit evaluations illuminates the structure of implicit social cognition. Even so, most investigations examine potential mechanisms in isolation, providing little insight on comparative or interactive effects of the basic processes contributing to implicit evaluation. Further, given the practical implications of addressing implicit racial preferences, it is surprising that there is little evidence about the relative effectiveness of different approaches. Paradigms have not been systematically compared while controlling for sample, setting, or procedural details that are irrelevant to the intervention itself. This presents a challenge for interpreting differences across paradigms. However, there is a solution: Comparative research can test complex interventions to identify *what* is effective, and—once identified—basic research can focus on understanding *why* the effective ones are effective. This division of labor may be more efficient than using mechanism-in-isolation research designs to identify both what is effective and why. Here, we pursue a research strategy that complements research focused on isolating mechanisms in order to understand how, when, and to what extent interventions are effective at changing the expression of implicit racial preferences.

Overview

We held a research contest to experimentally compare 17 interventions, a faking condition, and a control condition for reducing implicit racial preferences. We also investigated the interventions' effectiveness on explicit preferences and evaluations of other racial/ethnic groups (i.e., Hispanics and Asians). The contest took place over four studies. In the first study, we recruited researchers to submit interventions to reduce implicit preferences for Whites compared with Blacks. The presumed mechanisms underlying submitted interventions varied greatly, and included the following: increasing positive evaluations of Blacks, increasing negative evaluations of Whites, increasing control over the expression of biased racial attitudes, increasing self–outgroup overlap with Black individuals, and inducing egalitarian mindsets.

Between studies, teams could revise their interventions to be more effective, retain them as-is, or drop out of the contest. In total, 68 tests of implicit racial preference malleability were conducted; 17 interventions and a comparison condition designed to artificially induce malleability were tested an average of 3.78 times each. The findings provide evidence for differential effectiveness of approaches for reducing implicit racial preferences and set the stage for a new generation of research to clarify the mechanisms responsible for effective change.

Method

Participants

Participants in all studies were non-Black U.S. citizens/residents who registered at the Project Implicit research website (https://implicit.harvard.edu).[1] See Table 1 for sample characteristics. We report all data exclusions, conditions, and measures, and how we determined our sample size for each study. We had a simple decision rule for determining sample size in the first three studies: Data collection stops after each condition in the study had been assigned a set number of participants (300 in Studies 1 and 2, 400 in Study 3). For Study 4, we aimed to stop data collection once the study had reached 5,000 participants with completed sessions. On average, experimental conditions in all four studies had over 99% power to detect effects of a moderate effect size (Cohen's $d = .50$) and 61% power in Study 1, 64% power in Study 2, 42% power in Study 3, and 79% power in Study 4 to detect a small effect size (Cohen's $d = .20$).

Procedure

Participants volunteered for studies at Project Implicit's research site after completing a demographics registration form. Once registered, participants could visit the research website and be randomly assigned to studies from the research pool. Participants were assigned to the current studies only if they had never completed a study in the research pool. All studies are available for self-administration at https://osf.io/lw9e8/.

Studies 1 and 2. Participants were randomly assigned to complete a control condition, a faking comparison condition, or one of 13 intervention conditions in Study 1 and one of 14 intervention conditions in Study 2. In the control condition, participants did not complete an intervention task. Next, participants completed the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), followed by a self-report measure of racial attitudes.

Study 3. Study 3 was identical to Studies 1 and 2, except (a) there were 11 intervention conditions; (b) participants were randomly assigned to complete either the IAT or, a different implicit measure, the Multi-Category Implicit Association Test (Nosek, Sriram, Smith, & Bar-Anan, 2014) measuring evaluations of Whites, Blacks, Asians, and Hispanics simultaneously; and (c) participants completed additional self-report measures assessing attitudes toward Asian people and Hispanic people.

Study 4. There was evidence of differential attrition by condition in the first three studies.[2] Study 4 retained the 11 intervention conditions, control condition, and faking comparison condition in Study 3, and added a pretest IAT to test whether differential attrition could partially or fully explain results from the first three studies and examine change within subjects. However, completing a pretest may affect the posttest for some interventions and not others. We addressed this with a Solomon "four-group" design (Solomon, 1949) in which participants were randomly assigned to complete the pretest or not. This allows simultaneous examination of within-subjects change and potential testing-related effects on the posttest.

Dependent Measures

The IAT. The IAT assessed the relative strength of associations between two conceptual categories (i.e., White people, Black people) and two evaluative attributes (i.e., Good, Bad; see Nosek, Greenwald, & Banaji, 2007, for a review). The procedure followed the recommendations by Nosek, Greenwald, and Banaji (2005). Participants were instructed to categorize words and images as quickly as possible while also being accurate. The IAT is composed of seven blocks, with three practice blocks (omitted for analyses) and four critical blocks. In the first practice block (20 trials), participants categorized images of Black faces and White faces to categories labeled on the left or right. In the second practice block (20 trials), participants categorized good and bad words. In the third (20 trials) and fourth (40 trials) critical blocks, participants categorized images of Black faces/White faces and good/bad words on alternating trials. Consequently, participants categorized Black faces and bad words with one key and White faces and good words with another key. In the fifth practice block (40 trials), participants categorized images of White and Black faces again, except the categories had switched sides. The face category originally on the left was now categorized with the right key, and the face category originally on the right was now categorized with the left key. In the sixth (20 trials) and seventh (40 trials) critical blocks, participants categorized pairings opposite to the ones in the third and fourth blocks. Consequently, participants categorized White faces and bad words with one key and Black faces and bad words with the other key. The sixth and seventh blocks were counterbalanced with the third and fourth blocks between participants to control for potential order effects (Greenwald et al., 1998). The position of the Good/Bad categories was also randomized between participants: Half the participants categorized Good to the left key and Bad to the right key, whereas the other half did the reverse.

In Study 4, participants completed a shortened five-block IAT instead of the seven-block IAT to reduce study length. The five-block IAT was similar to the seven-block IAT with a few key changes. First, it had two combined categorization blocks instead of four and fewer trials (16 trials for the first two practice blocks, 32 trials for the combined blocks, and 24 trials for the practice block between the critical blocks).

The IAT was scored with the *D* algorithm recommended by Greenwald, Nosek, and Banaji (2003). A positive *D* score indicated faster responding on average when White faces were paired

[1] In all four studies, IAT scores were regressed on condition, race (White or non-White), and the interaction of condition and race. There was no interaction between condition and race in Study 1, $F(84, 3561) = 1.68$, $p = .05$, $\eta^2 = .0065$; Study 2, $F(85, 3977) = 1.36$, $p = .16$, $\eta^2 = .003$; Study 3, $F(82, 1973) = 1.32$, $p = .20$, $\eta^2 = .005$; or Study 4, $F(82, 4919) = 0.99$, $p = .50$, $\eta^2 = .02$.

[2] See the Appendix for descriptive statistics of attrition rate and http://osf.io/Lw9E8/ for the results of supplementary attrition analyses. We found evidence for differential attrition by condition in all four studies. There was scant evidence for experimental condition leading to differential attrition by demographics (i.e., age, religiosity [assessed on a 4-point Likert scale ranging from *Not at all religious* to *Very religious*]), gender, or political ideology (assessed on a 7-point Likert scale ranging from *Very Liberal* to *Very Conservative*). One exception was a statistically significant interaction of condition and age in predicting attrition in Study 3 ($p = .01$) that was not replicated across studies.

Table 1
Summary of Sample Characteristics

	N			Demographics of completed sessions		
Study	Began study	Completed study	Mean *N* / Condition	% Female	% White	Age *M*
Study 1	5,126	3,694	247	66.1	77.5	26.3
Study 2	5,581	4,111	257	65.3	75.3	26.7
Study 3	5,552	4,063	313	67.7	77.9	27.6
Study 4	7,732	5,116	394	64.0	77.2	31.3

with good words and Black faces were paired with bad words compared with the reverse. Positive scores are interpreted as an implicit preference for White people compared with Black people. *D* was calculated after removing response latencies under 400 ms or over 10,000 ms, and included all other trials. Categorization errors were replaced with the block mean of correct latencies plus 600 ms. Participants were excluded from the analyses if more than 10% of the critical response trials were faster than 300 ms, the error rate on any critical block was higher than 40%, or the overall error rate across all combined response blocks was over 30% (Nosek, Smyth, et al., 2007). For Study 4, participants were excluded from all IAT analyses if they met the exclusion criteria on either the pretest IAT or the posttest IAT. We excluded 116 (3.1%) participants in Study 1, 116 (2.8%) in Study 2, 47 (2.2%) in Study 3, and 292 (5.2%) in Study 4. Participant exclusion rates did not differ by condition in Study 1, $\chi^2(14, N = 3707) = 7.88$, $p = .90$; Study 2, $\chi^2(15, N = 4125) = 16.39$, $p = .36$; Study 3, $\chi^2(12, N = 2046) = 12.95$, $p = .37$; or Study 4, $\chi^2(12, N = 5604) = 16.56$, $p = .17$.

Multi-Category Implicit Association Test. For Study 3, about half the participants were randomly assigned to complete the race Multi-Category Implicit Association Test (MC-IAT; Nosek et al., 2014). The MC-IAT allowed us to investigate the degree to which implicit attitude malleability extended to another implicit measure and to preferences between Whites and other racial groups: Asians and Hispanics. The race MC-IAT differs from the race IAT by making two categories (instead of four) focal in each block of trials and, across a series of comparisons, providing comparative evaluations of four racial/ethnic categories instead of just two. Although the MC-IAT is also a relative measure, it has unique psychometric qualities and enables inferences about attitude change across multiple racial group comparisons.

The MC-IAT is composed of 14 blocks, with the first two blocks being practice blocks (omitted for analyses) and 12 critical blocks of 16 trials each. Participants responded by pressing *I* whenever a stimulus appeared in one of the two focal categories and by pressing *E* whenever a stimulus appeared that did not belong to a focal category. Each racial group was paired with good words as the focal category in three of the 12 critical blocks. The nonfocal categories were bad words and one of the other three racial groups. In total, participants completed three blocks each of the four possible focal categories: White people/Good, Black people/Good, Hispanic people/Good, and Asian people/Good. The participant pressed *I* whenever the target racial group or good word was presented and *E* whenever anything else was presented. Bad words were never a focal category. The three blocks for each target racial group had different racial groups as the nonfocal stimuli (e.g., the nonfocal racial group stimuli for the three White people/Good blocks rotated between Hispanic, Black, and Asian people).

The MC-IAT was computed using the *D* algorithm recommended by Nosek, Bar-Anan, Sriram, and Greenwald (2012). *D* was calculated after truncating response latencies under 200 ms to 200 ms and responses latencies over 2,000 ms to 2,000 ms. Participants were excluded from analyses if more than 10% of the responses in critical trials were under 200 ms. Fifty-two (2.6%) of completed MC-IATs were excluded on the basis of these criteria. IAT exclusions did not differ by condition, $\chi^2(12, N = 2007) = 7.19$, $p = .85$.

Self-reported racial attitudes. Participants completed three self-report items measuring racial attitudes in Studies 1–3. One assessed relative preference for White people over Black people on a 7-point Likert scale ranging from 1 (*I strongly prefer Black people to White people*) to 7 (*I strongly prefer White people to Black people*). The others were feeling thermometers for White people and Black people measured using a 7-point scale ranging from 1 (*Very cold*) to 7 (*Very warm*). In Study 3, participants also completed feeling thermometers for Asian people and Hispanic people. For analyses, a difference score was computed between the two feeling thermometers and averaged with the racial preference measure after standardizing each ($SD = 1$) while retaining a rational zero point of no preference between White people and Black people ($\alpha = .69, .69, .70$). Higher positive scores indicated a greater explicit preference for White people over Black people.

For Study 4, we tested the generality of interventions' effects on explicit racial attitudes with a 10-item version of the Subtle-Blatant Prejudice Scale adapted for prejudice toward Black people (Pettigrew & Meertens, 1995). The adapted scale contained five items from the Subtle Prejudice subscale and five items from the Blatant Prejudice subscale. Participants responded on 4-point Likert scales to statements. Higher scores indicated greater prejudice toward Black people.

Contest Qualification Criteria

A design incentive for participating researchers was to win the contest. To win, an intervention needed to elicit an average IAT score closest to the point of no implicit preference between Whites and Blacks and meet three qualification criteria. The qualification criteria were designed to maximize comparability on procedural elements, specifically intervention length, misbehavior on the IAT,

and attrition.[3] These criteria were established for winning the contest, not for determining whether the intervention was included in analyses. See the Appendix and an online supplement at https://osf.io/lw9e8/ for summaries of results on these criteria.

Background, Design, and Results

Mean IAT scores for each condition appear in Table 2. This section summarizes the effectiveness of interventions on reducing implicit preferences for White people compared with Black people across four studies. Following the intervention-by-intervention report, we summarize the comparative results across interventions, effects on explicit preferences, effects on the MC-IAT in Study 3, and effects related to the pretest IAT in Study 4. Positive t scores and Cohen's d effect sizes reflect larger reductions in preferences for White people over Black people relative to the control condition. For each intervention, we computed a fixed-effects meta-analytic effect size summarizing the aggregate effect of an intervention on implicit racial preferences. Participants in the control condition exhibited a moderate implicit preference for White people over Black people in all four studies (Ms = .45, .43, .50, .42; SDs = .39, .42, .43, .44, for Studies 1, 2, 3, and 4, respectively).

Interventions were organized into one of six descriptive categories highlighting features of the interventions.[4] The first category of interventions led participants to engage with others' perspectives (Interventions 1–3) by having participants imagine the thoughts, feelings, and actions of Black individuals. The second category exposed participants to counterstereotypical exemplars (Interventions 4–8); participants were assigned to fictional groups with positive Black ingroup members and/or negative White outgroup members or thought about famous Black people and infamous White people. The third category appealed to egalitarian values (Interventions 9–13) by activating egalitarian goals (e.g., thinking about failures to be objective or egalitarian) or having participants think about multicultural values. The fourth category used evaluative conditioning (Interventions 14 and 15) to strengthen counterstereotypical associations by pairing White faces with Bad words and Black faces with Good words. One approach in the fifth category (Intervention 16) attempted to reduce implicit preferences by inducing a positive emotion (elevation). Lastly, a sixth category reduced implicit preferences by providing strategies to override or suppress the influence of automatic biases, rather than trying to shift associations directly (Interventions 17 and 18).

Engaging With Others' Perspectives

Intervention 1: Training Empathic Responding (Hillary S. Schaefer)

Prejudice can manifest as a lack of empathy toward outgroup members (Avenanti, Sirigu, & Aglioti, 2010), and interventions designed to increase empathic responding can make explicit attitudes toward outgroup members more positive (Finlay & Stephan, 2000). To test whether empathy training can alter implicit preferences, participants played a game in which accurate empathic responding was rewarded. Participants observed a Black individual expressing an emotion (happy, sad, angry, or afraid). On each of 20 trials, they indicated which emotion was portrayed by picking from a list of four response options and then selected the likely reason the person was feeling this way from a list of four different scenarios (e.g., "I got a parking ticket" for anger). Participants were given positive feedback and points after each correct response. To maximize the personalization of the game, questions were asked in the first person ("What am I feeling?"), and correct feedback on the emotion rationale question was paired with a smiling face and the phrase "Thanks for understanding." In Study 1, empathy training did not elicit weaker implicit preferences compared with the control condition (M = .41, SD = .39), $t(513) = 1.06$, $p = .29$, $d = .10$.

In Study 1, participants took longer than the 5-min limit to complete the intervention (M = 5 min, 36 s). For Study 2, the intervention was revised to meet the time requirement by having two response options for each question instead of four. Fewer response options were also expected to make the task more engaging and effective. The revised empathy training task did not decrease implicit preferences (M = .47, SD = .41), $t(513) = 1.36$, $p = .17$, $d = -.10$. This intervention was not tested again in Studies 3 or 4. The meta-analytic effect size pooled from the two studies suggests that this empathy training game was ineffective at shifting implicit preferences ($d = -.02$, 95% CI [−.13, .10]).

Intervention 2: Perspective Taking (Steven A. Lehr)

Closeness with an outgroup member is correlated with less bias and weaker fear responses to outgroup faces (Olsson, Ebert, Banaji, & Phelps, 2005). Given the tendency to process close others in regions of the brain associated with processing the self (Mitchell, Banaji, & Macrae, 2005), reductions in implicit racial preferences may be moderated by associations with the self. Accordingly, this intervention aimed to activate self-referential processing in association with outgroup faces in Study 1. The intervention adapted a perspective-taking paradigm (Ames, Jenkins, Banaji, &

[3] First, at least 85% of participants had to complete the intervention in 5 min or less. This length is consistent with many published interventions to alter implicit associations. Many interventions did not meet the time requirement (M = 77.2% of participants completed within the defined time frame), but the average length of the intervention did not correlate reliably with reduced implicit preference for any of the studies (overall $r = -.09$), suggesting that time variation was not a biasing influence. All interventions were retained for comparative analysis. Second, the percentage of participants whose IAT performance was excluded had to be within two standard deviations of the mean exclusion rate across all interventions. This discouraged strategies that would compromise IAT interpretability (e.g., instructing participants to close their eyes and hit the keys randomly until the task is completed). However, this does not eliminate the possibility of faking strategies that meet inclusion criteria (see the Results section). Third, the percentage of participants who dropped out of the study prior to completing the IAT had to be within two standard deviations of the mean dropout rates across all interventions. This disincentivized designs that would induce attrition to select for individuals possessing lower preferences than the overall sample.

[4] These categorizations highlight the most prominent feature of the intervention design, but they do not unambiguously clarify the operative mechanisms. The most appropriate level of interpretation is the individual intervention itself. Nonetheless, there is considerable communication value in aggregating by dominant features. Complementary research that isolates operative mechanisms will clarify the appropriateness and limitations of the categorizations.

Table 2
Implicit Racial Preferences

Condition	Study 1			Study 2			Study 3			Study 4		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Control	274	.45	.39	302	.43	.42	182	.50	.43	462	.42	.44
Engaging with others' perspectives												
Training Empathic Responding	241	.41	.39	282	.47	.41						
Perspective Taking	232	.47	.39									
Imagining Interracial Contact	216	.44	.40	267	.43	.40						
Exposure to counterstereotypical exemplars												
Vivid Counterstereotypic Scenario	284	.35***	.44	241	.22***	.46	161	.17***	.46	373	.16***	.49
Practicing an IAT With Counterstereotypical Exemplars	274	.41	.42	229	.24***	.45	150	.32***	.43	399	.25***	.46
Shifting Group Boundaries Through Competition				279	.22***	.43	156	.26***	.45	392	.24***	.47
Shifting Group Affiliations Under Threat				284	.38	.42	166	.19***	.44	399	.32**	.46
Highlighting the Value of a Subgroup in Competition	294	.44	.42									
Appeals to egalitarian values												
Priming Feelings of Nonobjectivity	243	.45	.37	283	.47	.39	165	.46	.41	379	.38	.47
Considering Racial Injustice	235	.41	.43	252	.42	.42						
Instilling a Sense of Common Humanity				241	.46	.43	121	.45	.39	387	.42	.44
Priming an Egalitarian Mindset	245	.46	.37	233	.51*	.41	157	.43	.41	370	.39	.46
Priming Multiculturalism							146	.34***	.44	384	.31***	.46
Evaluative conditioning												
Evaluative Conditioning	117	.37	.41	205	.39	.38	166	.41*	.41	382	.29***	.47
Evaluative Conditioning With the GNAT	205	.38	.38	207	.25***	.43	116	.28***	.42	374	.29***	.49
Inducing emotion												
Inducing Moral Elevation	208	.45	.40	222	.37	.41						
Intentional strategies to overcome biases												
Using Implementation Intentions	249	.37*	.44	235	.30***	.40	152	.31***	.43	376	.17***	.52
Faking the IAT	274	.37*	.51	247	.21***	.60	161	.25***	.58	344	.15***	.67

Note. N = number of completed Implict Association Tests (IATs) for the condition. IAT means are D scores (Greenwald et al., 2003), and positive values indicate greater preference for White people compared with Black people. GNAT = go/no-go association task.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Mitchell, 2008) that showed Black faces accompanied by an emotional context (e.g., "This person just found a $100 bill on the ground"). Participants imagined that they were the person in the situation and wrote about how they felt. Participants provided short statements for five situations. This intervention was ineffective at reducing implicit preferences ($M = .47$, $SD = .39$), $t(504) = .45$, $p = .66$, $d = -.04$, 95% CI [−.22, .14],[5] and was not tested again in any of the other studies.

Intervention 3: Imagining Interracial Contact (Eva E. Chen and Rhiannon N. Turner)

The contact hypothesis states that, under the right conditions, contact between members of different groups will lead to more positive intergroup relations (Allport, 1954). Mental imagery can elicit similar emotional and motivational responses as real experiences (Dadds, Bovbjerg, Redd, & Cutmore, 1997), suggesting that imagining intergroup contact might have similar consequences to the actual experience. Supporting this idea, participants who imagine contact with a range of outgroups (e.g., older adults, Muslims, gay people) subsequently show less intergroup anxiety and hold more positive attitudes toward that group (Crisp & Turner, 2009; Turner & Crisp, 2010; Turner, Crisp, & Lambert, 2007). In this intervention, participants imagined interacting with a Black stranger in a relaxed, positive, and comfortable environment. Then they listed as many details as possible about the imagined interaction. This intervention was ineffective in Study 1 ($M = .44$, $SD = .40$), $t(488) = .42$, $p = .68$, $d = .03$.

For Study 2, participants were instructed not only to imagine a positive interaction with a Black person but also to imagine a negative interaction with a White person. Along with the corresponding prompts, participants saw a photograph of a smiling Black woman and a photograph of a frowning White woman. The purpose of these changes was to increase participants' positive thoughts about Black people by providing a contrasting negative encounter with a White person, and to make the encounters more vivid by providing specific images of the strangers. The revised intervention did not decrease implicit preferences ($M = .43$, $SD = .40$), $t(567) = .12$, $p = .90$, $d = 0.00$, and was not tested again in Studies 3 or 4.

In contrast to prior research showing that intergroup contact can reduce implicit preferences (Turner & Crisp, 2010), the meta-analytic effect size suggests that imagining intergroup contact with a Black person did not decrease implicit racial preference across two experiments ($d = .01$, 95% CI [−.11, .13]).

[5] Reported confidence intervals reflect confidence intervals of Cohen's d.

Exposure to Counterstereotypical Exemplars

Intervention 4: Vivid Counterstereotypic Scenario (Maddalena Marini, Sandro Rubichi, and Giuseppe Sartori)

Foroni and Mayr (2005) shifted implicit preferences for flowers versus insects by first presenting a fictional counterstereotypic scenario in which flowers were dangerous and insects were good. Furthermore, personal relevance is known to increase processing of persuasive messages (for a review and meta-analysis, see Johnson & Eagly, 1989). Putting the participant "into" the story through second-person narratives might increase the story's impact. In Study 1, participants read an evocative story told in second-person narrative in which a White man assaults the participant and a Black man rescues the participant. Participants were also told that the task following the story (i.e., the race IAT) was supposed to affirm the associations: White = Bad, Black = Good. Participants were instructed to keep the story in mind during the IAT. This intervention successfully reduced implicit preferences ($M = .35$, $SD = .44$), $t(556) = 2.93$, $p = .0035$, $d = .24$. In Study 2, length and vividness of the story were increased (e.g., from "With sadistic pleasure, he bashes you with his bat again and again" to "With sadistic pleasure, he beats you again and again. First to the body, then to the head. You fight to keep your eyes open and your hands up. The last things you remember are the faint smells of alcohol and chewing tobacco and his wicked grin"). This intervention was more than doubly effective in reducing implicit preferences than in Study 1 ($M = .22$, $SD = .46$), $t(541) = 5.54$, $p = 4.61 \times 10^{-8}$, $d = .48$.

For Study 3, the instructions to affirm positive Black associations and negative White associations were revised to include two sets of pictures. One set shows the stimuli for Black people on the IAT and MC-IAT paired with the word *good*, whereas the other set showed the stimuli for White people on the IAT and MC-IAT paired with the word *bad*. Participants exhibited decreased implicit preferences ($M = .17$, $SD = .46$), $t(346) = 6.80$, $p = 2.17 \times 10^{-11}$, $d = .75$. Study 4's intervention was revised to only include one set of pictures, as the MC-IAT was not used. Implicit preferences were significantly weaker than control ($M = .16$, $SD = .49$), $t_{sattherwaite}(755.24) = 8.06$, $p = 3.04 \times 10^{-15}$, $d = .57$. Across four studies, the meta-analytic effect size suggests that the vivid, second-person counterstereotypic scenario substantially reduces implicit preferences ($d = .49$, 95% CI [.41, .58]).

Intervention 5: Practicing an IAT With Counterstereotypical Exemplars (Bethany A. Teachman)

A variation of the IAT procedure was used in this intervention to reinforce positive associations with Blacks and negative associations with Whites. Participants repeatedly practiced the combined response blocks of the race IAT that paired Black with Good and White with Bad. The reverse pairing associating Blacks with Bad did not appear during the intervention. The stimulus items representing Blacks and Whites were the same as those used in the race IAT, plus six positive, well-known Black exemplars (e.g., Bill Cosby) and six negative White exemplars (e.g., Charles Manson). Prior research demonstrates that exposure to positive Black and negative White exemplars can shift implicit racial preferences (Dasgupta & Greenwald, 2001; Joy-Gaba & Nosek, 2010). Also, past evidence of order effects based on the sequence of the combined response blocks on the IAT (Nosek et al., 2005) suggest that the IAT can serve as an intervention itself, at least temporarily shifting associations. The combination of the positive Black and negative White practice was anticipated to be particularly effective.

Due to a programming error in Study 1, participants learned that they were going to perform part of a race IAT and saw the positive Black and negative White exemplars that would accompany the standard Black and White images, but they did not actually complete the counterstereotypic practice. This manipulation did not reduce implicit preferences ($M = .41$, $SD = .42$), $t(546) = 1.04$, $p = .30$, $d = .10$. Because of the error, this result was not used in later analyses aggregating the results from this intervention.

For Study 2, the procedure was implemented as described above. This intervention reduced implicit preferences ($M = .24$, $SD = .45$), $t(529) = 4.85$, $p = .0000016$, $d = .44$. For Study 3, the counterstereotypic practice was reduced from 90 trials to 52 to align with time requirements. This intervention reduced implicit preferences to the same degree ($M = .32$, $SD = .43$), $t(330) = 3.87$, $p = .00013$, $d = .43$. Study 4 retained the same design and the intervention also reduced implicit preferences ($M = .25$, $SD = .46$), $t(859) = 5.45$, $p = 6.75 \times 10^{-8}$, $d = .37$. Combining Studies 2–4, the meta-analytic effect size suggests that this intervention was effective at reducing implicit preferences ($d = .40$, 95% CI [.30, .49]).

Intervention 6: Shifting Group Boundaries Through Competition (Rebecca S. Frazier)

Intense competition and strong outgroup threats lead to negative outgroup attitudes (Riek, Mania, & Gaertner, 2006). Cooperating with racial outgroup members to compete against White ingroup members may decrease implicit preferences for White people over Black people. In this intervention, participants were assigned to be part of a dodgeball team in which all of their teammates were Black and all of their opponents were White. Participants targeted White opponents and received aid from their Black teammates during play. Members of the opposing all-White team engaged in unfair play. At the end of the intervention, participants made goal intentions to think "good = Black" and "bad = White" and to remember how their Black teammates helped them and their White enemies hurt them. This intervention was not tested in Study 1. In Study 2, the intervention was successful in reducing implicit preferences ($M = .22$, $SD = .43$), $t(579) = 5.77$, $p = 1.28 \times 10^{-8}$, $d = .50$. To adhere with contest time requirements in Study 3, sections requiring participant input were set to automatically advance if participants responded too slowly. This intervention reduced implicit preferences ($M = .23$, $SD = .53$), $t(336) = 5.11$, $p = 5.53 \times 10^{-7}$, $d = .56$. Study 4's intervention retained the same design and replicated Study 3's result ($M = .24$, $SD = .47$), $t(852) = 5.78$, $p = 1.03 \times 10^{-8}$, $d = .40$. Overall, the meta-analytic effect size suggests that this intervention was successful in reducing implicit preferences ($d = .45$, 95% CI [.36, .55]).

Intervention 7: Shifting Group Affiliations Under Threat (Steven A. Lehr)

Under conditions of heightened threat, people may become more attentive to cues of coalition membership, because sharpened differentiation between friends and foes would be adaptive when one is endangered. Accordingly, participants in this intervention read a vivid postapocalyptic scenario that was highly threatening. In Study 2, participants were then shown faces of "friends," most of whom were Black, along with descriptions suggesting that they might be valuable as alliances (e.g., highlighting a medical background or hunting abilities). This intervention was not administered in Study 1. In Study 2, this intervention did not decrease implicit preferences ($M = .38$, $SD = .42$), $t(584) = 1.40$, $p = .16$, $d = .12$. Past research suggests that exposure to only positive Black figures may be less effective at changing implicit racial attitudes than exposure to both positive Black and negative White exemplars (Joy-Gaba & Nosek, 2010). The same paradigm from Study 2 was used in Study 3, but faces and descriptions of "enemies" who were all White people were added. Including a negative White contrast group was expected to make the ingroup–outgroup distinction more salient. These changes led to decreased implicit preferences ($M = .19$, $SD = .44$), $t(346) = 6.80$, $p = 4.57 \times 10^{-11}$, $d = .73$. Study 4 retained the same design but changed the faces of the Black individuals to be more likable and the faces of the White individuals to be less likable. Study 4's intervention was effective, but was less effective than Study 3's ($M = .33$, $SD = .46$), $t(859) = 3.12$, $p = .002$, $d = .21$. Overall, the meta-analytic effect size suggests that this intervention was successful in reducing implicit preferences ($d = .28$, 95% CI [.18, .37]).

Intervention 8: Highlighting the Value of a Subgroup in Competition (Selin Kesebir)

The common ingroup identity model anticipates that emphasizing superordinate identities may reduce biases toward outgroup members (Gaertner & Dovidio, 2000). Reminding people that African Americans have contributed to America's international standing may highlight a superordinate group identity (American) that includes African Americans and make positive contributions of African Americans more accessible. Participants read a description of international competition in basketball that stated that the United States has one of the most successful basketball teams in the world, but is now facing heavy competition from other countries. Participants were then presented with a list of eight prominent basketball players' names (i.e., Dwyane Wade, Kobe Bryant, Jason Terry, Steve Nash, Brent Barry, Tim Duncan, Shaquille O'Neal, Kevin Garnett) and asked to mark which ones they recognized. This questionnaire aimed to indirectly remind participants of the mostly Black demographic composition of American basketball, though the racial identity of individual players was not made explicit. In Study 1, this intervention was ineffective at reducing implicit preferences ($M = .44$, $SD = .42$), $t(566) = .32$, $p = .75$, $d = .03$, 95% CI [−.14, .19], and was not tested again in any of the other studies.

Appeals to Egalitarian Values

Intervention 9: Priming Feelings of Nonobjectivity (Jennifer A. Joy-Gaba)

The phenomenon known as the "bias blind spot" (Pronin & Kugler, 2007) reflects people's beliefs that they are objective and impartial and are immune to biased judgments. For example, Uhlmann and Cohen (2007) demonstrated that having individuals affirm their objectivity ironically leads to more discrimination. Perhaps inducing feelings of *non*objectivity would cue people to take control over their automatic biases and decrease subsequent racial preferences. Study 1 attempted to induce these feelings using Schwarz and colleagues' (1991) ease-of-retrieval paradigm, in which the difficulty of remembering examples affects how much individuals perceive that they possess the characteristic (Schwarz, 1998). In Study 1, participants attempted to recall nine past examples in which they behaved objectively, presuming that the difficulty in generating examples would induce self-doubt about the ability to act objectively. This intervention was ineffective at reducing implicit preferences ($M = .45$, $SD = .37$), $t(515) = .07$, $p = .95$, $d = 0.00$.

For Study 2, the conceptual hypothesis was retained but the intervention was changed to use Devine, Monteith, Zuwerink, and Elliot's (1991) should-would discrepancy paradigm. Here, participants reported how they personally *would* act given a particular decision compared with how society believes they *should* act when making a decision. For example, individuals might report that there *would* be times when they would make a choice based solely on their preference, without consideration of the facts, while simultaneously reporting that they *should* consider all the facts when making a decision. Reporting a discrepancy between participants' "would" responses and their "should" responses may lead to self-awareness of nonobjectivity, which would cue people to take control over their automatic biases. This alternative also failed to decrease implicit preferences ($M = .47$, $SD = .39$), $t(583) = 1.29$, $p = .20$, $d = -.11$. For Study 3, participants read a fictitious excerpt from a popular science article about psychological biases outside of conscious awareness that may influence behavior (adapted from Pronin & Kugler, 2007). This approach to priming nonobjectivity was unsuccessful at decreasing preferences ($M = .46$, $SD = .41$), $t(345) = 1.06$, $p = .29$, $d = .08$. Study 4's intervention retained the same design and was similarly ineffective at decreasing preferences ($M = .38$, $SD = .47$), $t(839) = 1.31$, $p = .19$, $d = .09$. Across four experiments, the meta-analytic effect size suggests that priming feelings that one is nonobjective did not decrease preferences ($d = -.02$, 95% CI [−.06, .10]).

Intervention 10: Considering Racial Injustice (Larisa Heiphetz)

Many Americans believe that racial inequality is a thing of the past (Eibach & Ehrlinger, 2006), that the current social order is legitimate (Jost & Banaji, 1994), and that Whites deserve the benefits they receive (Bonilla-Silva, Lewis, & Embrick, 2004). These perceptions may encourage people to view Whites as members of an admirable group and contribute to implicit preferences for Whites over Blacks. In this intervention, participants listed two examples of injustices that Whites inflicted on Blacks in the past,

two examples of injustices that Whites currently inflict on Blacks, and two examples of ways in which Blacks have overcome racial injustice. Considering harms perpetuated by Whites could cause participants to view this group less positively, whereas thinking about Blacks' efforts to overcome inequality could lead to perceptions of Blacks as agents of positive social change and thus make attitudes toward Blacks more favorable. In Study 1, considering racial injustices did not significantly reduce implicit preferences ($M = .41$, $SD = .43$), $t(507) = 1.03$, $p = .31$, $d = .10$. In Study 2, participants wrote about one example in each category rather than two to make it easier and shorter. However, considering racial injustices did not reduce implicit preferences in Study 2 ($M = .42$, $SD = .42$), $t(552) = .04$, $p = .97$, $d = .02$. This intervention was not tested in Studies 3 or 4. The meta-analytic effect size suggests that considering racial injustices was not effective at reducing implicit preferences across two studies ($d = .05$, 95% CI [−.08, .17]).

Intervention 11: Instilling a Sense of Common Humanity (Carlee Beth Hawkins)

Social attitudes are influenced by the boundaries drawn between ingroups and outgroups, with ingroup members being liked more than outgroup members (Tajfel & Turner, 1979). If the boundaries of the ingroup can be redrawn and expanded to include outgroup members, attitudes toward outgroups may become more positive (Gaertner, Dovidio, Anastasio, Bachman, & Rust, 1993). To test this, participants viewed a popular video clip of a man dancing with people in different countries all over the world (http://www.youtube.com/watch?v=zlfKdbWwruY). The video communicates a sense of common humanity, demonstrates that all types of people can be happy and silly together, and may redefine boundaries of the ingroup. This feeling of common humanity was predicted to create more positive attitudes toward racial outgroups. This intervention was not tested in Study 1. In Study 2, this intervention did not decrease implicit preferences ($M = .46$, $SD = .43$), $t(541) = 1.07$, $p = .29$, $d = -.07$. This intervention was tested without revision in Study 3. Again, the intervention did not decrease implicit preferences ($M = .45$, $SD = .39$), $t(301) = 1.21$, $p = .23$, $d = .14$. The intervention was tested a third time in Study 4 and did not reveal evidence for decreased implicit preferences ($M = .42$, $SD = .44$), $t(847) = .14$, $p = .89$, $d = .01$. Across three experiments, the meta-analytic effect size suggests that this method of instilling a sense of common humanity did not decrease implicit preferences ($d = .00$, 95% CI [−.10, .10]).

Intervention 12: Priming an Egalitarian Mindset (Arnold K. Ho)

Priming social ideologies can shift racial preferences (e.g., Katz & Hass, 1988; Sears & Henry, 2005; Sidanius & Pratto, 1999). Katz and Hass (1988) primed egalitarianism with the Humanitarian-Egalitarianism scale and found that it increased explicit pro-Black attitudes, especially among participants who scored above the median on the scale. This design was replicated with the expectation that egalitarianism priming would attenuate the implicit preferences. This intervention did not decrease implicit preferences in Study 1 ($M = .46$, $SD = .37$), $t(517) = .30$, $p = .77$, $d = -.03$.[6] For Study 2, participants wrote a short essay in favor of the statement, "All people and groups are equal; therefore, they should be treated the same way." Unexpectedly, this manipulation increased implicit preferences ($M = .51$, $SD = .41$), $t(533) = -2.24$, $p = .026$, $d = -.19$. For Study 3, the operationalization of the hypothesis was changed again. In an effort to prime egalitarian goals as opposed to values alone (Moskowitz & Li, 2011), participants were given a questionnaire that asked them how important it was to be egalitarian. After, participants wrote about a time they failed to live up to egalitarian ideals. Activating egalitarian goals was predicted to decrease implicit preferences, but this intervention did not decrease implicit preferences ($M = .43$, $SD = .41$), $t(337) = 1.67$, $p = .10$, $d = .18$. This intervention was retained without revision in Study 4 and did not decrease implicit preferences ($M = .39$, $SD = .46$), $t(830) = .84$, $p = .40$, $d = .06$. Overall, the meta-analytic effect size suggests that priming an egalitarian mindset was ineffective in decreasing implicit preferences ($d = .00$, 95% CI [−.09, .08]).

Intervention 13: Priming Multiculturalism (Larisa Heiphetz)

To reduce racial animus, some advocate colorblindness—the idea that racial categories are unimportant and should not be taken into account when crafting public policy. However, experimental evidence suggests that color blindness elicits stronger racial preferences compared with multiculturalism—the idea that racial differences should be acknowledged and celebrated (Richeson & Nussbaum, 2004). This intervention examined the effect of multiculturalism on racial attitudes by encouraging participants to adopt a multicultural perspective. Following Richeson and Nussbaum (2004), participants read a prompt advocating multiculturalism, summarized the prompt in their own words, and then listed two reasons why multiculturalism "is a positive approach to interethnic relations." Finally, participants were given instructions to think "Black = Good" as they took the IAT/MC-IAT. This intervention was not tested in Studies 1 and 2. In Study 3, this intervention decreased implicit preferences ($M = .34$, $SD = .44$), $t(326) = 3.34$, $p = .00094$, $d = .37$. Study 4's intervention retained the same design and also decreased implicit preferences ($M = .31$, $SD = .46$), $t(844) = 3.7$, $p = 2.30 \times 10^{-4}$, $d = .26$. The overall meta-analytic effect size across two experiments suggests that priming multiculturalism reduced implicit racial preferences ($d = .29$, 95% CI [.17, .40]).

Evaluative Conditioning

Intervention 14: Evaluative Conditioning (Sean P. Wojcik and Spassena P. Koleva)

Evaluative conditioning is the process by which pairing an attitude object with another valenced attitude object shifts attitudes of the first object in the direction of the valenced object (De Houwer, Thomas, & Baeyens, 2001; Olson & Fazio, 2001, 2002,

[6] This intervention replicated the original demonstration of reducing explicit racial preferences ($M = .42$, $SD = .74$), $t(516) = 2.37$, $p = .018$, $d = 21$.

2006). In this intervention, participants saw pairings of Black faces with positive words and White faces with negative words. Participants viewed 48 pairs of a face image and a valenced word one at a time in the center of the screen. The stimuli were the face stimuli and words from the race IAT. African American faces were always paired with positive words, and White faces were always paired with negative words. In each trial, the stimuli appeared together for 1 s. After presentation of the stimuli, participants categorized the face and memorized the valence words for later testing. Participants pressed either the *E* or *I* key to indicate each face's race, and the correct key response was randomized for each trial. After the categorization task, participants recalled as many of the valence words as possible. In Study 1, this intervention did not decrease implicit preferences ($M = .37$, $SD = .41$), $t(389) = 1.81$, $p = .071$, $d = .19$. Due to a programming error, *Internet Explorer* users were unable to complete the intervention in Study 1. This error was fixed in Study 2. To reduce task difficulty, participants just read the words (instead of memorizing them) as they categorized faces. The revised intervention did not decrease implicit preferences ($M = .39$, $SD = .38$), $t(505) = .92$, $p = .36$, $d = .10$. For Study 3, the instructions were changed back to those from Study 1. Participants memorized the words as they were presented on the screen. The number of trials was also reduced from 48 to 40 to make it shorter. This intervention reduced implicit preferences ($M = .42$, $SD = .41$), $t(346) = 1.99$, $p = .05$, $d = .21$. Study 4's intervention retained the same design and reduced implicit preferences ($M = .29$, $SD = .47$), $t(842) = 4.09$, $p = 4.71 \times 10^{-5}$, $d = .28$. Overall, the meta-analytic effect size suggests that this evaluative conditioning task was effective at reducing implicit preferences ($d = .21$, 95% CI [.12, .30]).

Intervention 15: Evaluative Conditioning With the Go/No-Go Association Task (Carlo Cerruti and Jiyun-Elizabeth L. Shin)

An adapted go/no-go association task (GNAT; Nosek & Banaji, 2001), which is characterized by rapid and repeated associations between two stimuli, was used to strengthen the association between Black people and "good." In this GNAT, participants responded to stimuli pairings when they fulfilled two categories ("go") and abstained from responding when the stimuli pairings did not fit the categories ("no-go"). In the first block, participants were instructed to "go" when the stimulus pairing was a Black person and a Good word, and to do nothing when the stimulus pairing was not a Black person and a Good word. The majority of stimulus pairings were composed of Black people and Good words. In the second block, participants were instructed to "go" when the stimulus pairing was a White person and a Good word, and to do nothing if it was not. A minority of trials contained a White person and Good word pairing. Thus, the second block attempted to discourage automatic associations of White people and Good words. In Study 1, this intervention did not reduce implicit preferences ($M = .38$, $SD = .38$), $t(477) = 1.90$, $p = .059$, $d = .18$.

The following changes were made for Study 2: (a) The total number of trials was reduced from 100 to 60; (b) both blocks used the same "go" category; the "go" category for the second block was "Black and Good" instead of "White and Good"; and (c) the second block required faster responses than the first, and participants were encouraged on an instructions screen to respond faster. These changes enhanced the effectiveness of the intervention in decreasing implicit preferences ($M = .25$, $SD = .43$), $t(507) = 4.71$, $p = .0000031$, $d = .54$. In Study 3, participants were instructed to count the number of times Black faces and Good words were categorized together in the task. This change was expected to strengthen the conditioning effect. Also, the number of trials was reduced from 60 to 45 to shorten the task. The intervention successfully replicated the effects of Study 2; the intervention decreased implicit preferences ($M = .28$, $SD = .42$), $t(296) = 4.51$, $p = .0000092$, $d = .45$. Study 4's evaluative conditioning task retained the same design and also decreased implicit preferences ($M = .29$, $SD = .49$), $t_{sattherwaite}(754.92) = 3.95$, $p = .000085$, $d = .28$. Overall, the meta-analytic effect size shows that evaluative conditioning with the GNAT decreased implicit preferences ($d = .32$, 95% CI [.24, .41]).

Inducing Emotion

Intervention 16: Inducing Moral Elevation (Jonathan Haidt)

The emotion of "elevation" (Algoe & Haidt, 2009; Haidt, 2003) is induced from witnessing acts of charity, gratitude, or generosity. We hypothesized that elevation would blur boundaries between the ingroup and the outgroup and, consequently, lead to weaker implicit racial preference for Whites compared with Blacks. Participants watched an elevating video about a high school girls' softball game in which players showed extraordinary sportsmanship by carrying an opposing player around the bases after she injured herself as she hit a homerun. No Black people was present in the video, and prior research confirmed that this video induced moral elevation (Lai, Haidt, & Nosek, 2013). In Study 1, elevation was not effective at reducing implicit preferences ($M = .45$, $SD = .40$), $t(480) = .10$, $p = .92$, $d = 0.00$. Perhaps, moral elevation induced by Black individuals could be more effective at reducing implicit racial preferences. For Study 2, the video was changed to one that showed Black people behaving in elevating ways. In the video, a Black high school music teacher expresses his gratitude toward his former music teacher (also Black), who had seen promise in the young man when he was a teenager and saved him from a life of crime. Pretesting confirmed that this video induced moral elevation. In Study 2, this moral elevation induction was ineffective in reducing implicit preferences ($M = .37$, $SD = .41$), $t(522) = 1.41$, $p = .16$, $d = .15$. This intervention was not tested in Studies 3 or 4. Overall, the meta-analytic effect size suggests that moral elevation was ineffective in reducing implicit preferences ($d = .06$, 95% CI [−.06, .19]).

Intentional Strategies to Overcome Biases

Intervention 17: Using Implementation Intentions (Calvin K. Lai)

Implementation intentions are if-then plans that tie a behavioral response to a situational cue (Gollwitzer, 1999). Setting implementation plans are effective at increasing the consistency between goal-directed intentions and behavior by increasing behavioral

automaticity. The mechanism connects an environmental cue with the goal intention, making associations between the behavior and the cue more accessible in memory (Brandstätter, Lengfelder, & Gollwitzer, 2001). Stewart and Payne's (2008) implementation intentions manipulation was adapted for the current intervention. The task gave participants a short tutorial on how to take the IAT and informed them about the tendency for people to exhibit an implicit preference for Whites compared with Blacks. Participants were then asked to commit themselves to an implementation intention by saying to themselves silently, "I definitely want to respond to the Black face by thinking 'good.'" In Study 1, making implementation intentions decreased implicit preferences (M = .37, SD = .44), $t(521) = 2.15$, $p = .032$, $d = .19$. In Study 2, participants completed practice trials of the IAT to familiarize them with the task before being given the implementation intention instructions. Intervention effectiveness was increased compared with Study 1; participants in the implementation intentions condition exhibited decreased implicit preferences (M = .30, SD = .40), $t(535) = 3.58$, $p = .00037$, $d = .32$. In Study 3, the intervention instructions were revised to include information about the MC-IAT. Participants still completed a practice IAT before instructions (not an MC-IAT). The intervention decreased implicit racial preferences (M = .31, SD = .43), $t(332) = 4.16$, p = .000041, d = .46. This manipulation was retained without revision in Study 4 and also decreased IAT scores (M = .17, SD = .52), $t_{sattherwaite}(731.05) = 7.38$, $p = 4.25 \times 10^{-13}$, d = .52. In line with prior research (Gollwitzer & Schaal, 1998; Stewart & Payne, 2008), implementation intentions were effective in decreasing implicit preferences (d = .38, 95% CI [.30, .47]).

Intervention 18: Faking the IAT (Calvin K. Lai)

Although the IAT is relatively resistant to faking (Banse, Seise, & Zerbes, 2001; Kim, 2003; Steffens, 2004), it can be manipulated through the use of behavioral strategies (Fiedler & Bluemke, 2005). As a comparison condition to the "real" interventions, participants completed a modified version of Cvencek, Greenwald, Brown, Gray, and Snowden's (2010) faking manipulation. This provided an opportunity to observe whether actual interventions are distinguishable from faking effects. Participants were given a short tutorial on how to take the IAT and informed about the tendency for people to exhibit an implicit preference for Whites compared with Blacks. Participants were then told the study was about faking the IAT and were asked to slow down on blocks with "Black and Bad" paired together and speed up on blocks with "White and Bad" paired together. None of the interventions could change the instructions for the IAT used as the dependent variable, and the IAT instructions encouraged "accurate" behavior by participants. The faking manipulation anticipated this by instructing participants to ignore instructions for the IAT that contradicted the faking instructions. The purpose of including this manipulation was to obtain comparative data with an intervention that is likely to be manipulating task performance rather than changing associations or increasing control over the expression of those associations. The comparative insights are addressed in the General Discussion.

Faking successfully reduced IAT scores in Study 1 (M = .37, SD = .51), $t_{sattherwaite}(513.67) = 1.97$, $p = .047$, $d = .18$. In Study 2, to familiarize participants with the task, IAT practice trials were presented before being given faking instructions. Effectiveness was increased compared with Study 1; participants in the faking condition exhibited lower IAT scores (M = .21, SD = .60), $t_{sattherwaite}(422.42) = 4.78$, $p = .0000025$, $d = .47$. In Study 3, the instructions were revised to include information about faking the MC-IAT. Participants still completed a practice IAT before instructions (not an MC-IAT). This manipulation decreased IAT scores (M = .25, SD = .58), $t_{sattherwaite}(289.91) = 4.57$, p = .0000072, d = .51. This manipulation was retained without revision in Study 4 and also decreased IAT scores (M = .16, SD = .67), $t_{sattherwaite}(555.52) = 6.25$, $p = 8.10 \times 10^{-10}$, d = .51. Overall, the meta-analytic effect size suggests that faking the IAT was effective at reducing IAT scores (d = .39, 95% CI [.31, .47]).

Implicit–Explicit Relations

Comparing individuals across all conditions in the current studies, implicit and explicit racial preferences were positively related: Study 1, $r(3421) = .23$; Study 2, $r(3791) = .24$; Study 3, $r(1885) = .22$; Study 4, $r(4946) = .22$. To test the effects of interventions on the strength of implicit–explicit relations, we constructed five general linear models (one for each implicit measure used in the four studies), with condition, implicit racial preferences, and the interaction between condition and implicit racial preferences predicting explicit racial preferences. We found main effects of implicit racial preferences on explicit preferences in all studies: Study 1, $F(1, 3393) = 179.47$, $p < 1 \times 10^{-36}$, η^2 = .050; $F(1, 3759) = 251.71$, $p < 1 \times 10^{-36}$, $\eta^2 = .063$; Study 3 IAT, $F(1, 1859) = 90.55$, $p < 5.38 \times 10^{-21}$, $\eta^2 = .046$; Study 3 MC-IAT, $F(1, 1763) = 54.55$, $p = 2.33 \times 10^{-13}$, $\eta^2 = .030$; Study 4, $F(1, 4922) = 244.38$, $p < 1 \times 10^{-36}$, $\eta^2 = .047$. In Studies 2 and 4, we found main effects of condition on explicit preferences, $F(15, 3759) = 2.12$, $p = .0070$, $\eta^2 = .008$; $F(12, 4922) = 251.71$, $p = .011$, $\eta^2 = .005$. In Study 2 only, we also found an interaction between condition and implicit preferences, $F(15, 3759) = 2.41$, $p = .0017$, $\eta^2 = .010$. In general, these findings suggest that the interventions did not alter the strength of implicit–explicit relations. See Table 3 for a summary of implicit–explicit correlations by condition.

Explicit Racial Preferences

The focus of the research contest is on reducing implicit preferences. However, it is of theoretical and practical interest to also understand what interventions are effective at changing explicit preferences. Overall, participants self-reported preferences for Whites over Blacks in Studies 1–3 (Ns = 3532, 3899, 3793; Ms = .45, .50, .50; SDs = .89, .88, .88), a moderate effect size (ds = .51–.57). Participants in Study 4 tended to hold nonprejudiced atttiudes toward Blacks on the Subtle-Blatant Prejudice scale in Study 4 (N = 5098, M = 1.76, SD = .44; range = 1–4).To examine the effects of interventions on explicit racial preferences, we modeled explicit preferences as a function of condition for each of the three studies. One-way analyses of variance revealed that there were no significant effects of condition on explicit racial preferences in Study 1, $F(14, 3517) = 1.08$, $p = .38$, $\eta^2 = .004$; Study 2, $F(15, 3883) = 1.43$, $p = .12$, $\eta^2 = .006$; Study 3, $F(12, 3780) = 1.05$, $p = .40$, $\eta^2 = .003$; or Study 4, $F(12, 5085) = 1.28$, $p = .23$, $\eta^2 = .003$. See Table 3 for a summary of explicit racial

Table 3
Explicit Racial Preferences

Condition	Study 1			Study 2			Study 3			Study 4		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Control	272	.58	.80	291	.53	.96	338	.49	.93	473	1.76	.46
Engaging with others' perspectives												
Training Empathic Responding	233	.45	.92	274	.45	.83						
Perspective Taking	229	.55	.93									
Imagining Interracial Contact	207	.49	.91	248	.47	.87						
Exposure to counterstereotypical exemplars												
Vivid Counterstereotypic Scenario	282	.51	.87	247	.38	.91	319	.50	.94	375	1.73	.41
Practicing an IAT with Counterstereotypical Exemplars	267	.47	.82	224	.33*	.83	292	.54	.91	410	1.72	.42
Shifting Group Boundaries Through Competition				277	.51	.90	287	.49	.82	402	1.78	.44
Shifting Group Affiliations Under Threat				282	.36*	.80	324	.46	.84	401	1.74	.44
Highlighting the Value of a Subgroup in Competition	296	.48	.97									
Appeals to egalitarian values												
Priming Feelings of Nonobjectivity	238	.53	.86	279	.53	.88	323	.50	.94	382	1.76	.46
Considering Racial Injustice	229	.42*	.86	234	.49	.83						
Instilling a Sense of Common Humanity				245	.36*	.79	240	.57	.84	393	1.77	.42
Priming an Egalitarian Mindset	246	.41*	.74	209	.47	.96	267	.58	.91	378	1.76	.41
Priming Multiculturalism							269	.40	.79	388	1.75	.43
Evaluative conditioning												
Evaluative Conditioning	115	.38*	.82	202	.54	.98	290	.44	.88	382	1.74	.43
Evaluative Conditioning With the GNAT	199	.46	.96	201	.51	.75	242	.48	.86	387	1.76	.47
Inducing emotion												
Inducing Moral Elevation	207	.51	.94	216	.45	1.06						
Intentional strategies to overcome biases												
Using Implementation Intentions	244	.59	.89	228	.44	.86	294	.45	.81	374	1.74	.43
Faking the IAT	268	.57	.97	242	.49	.84	308	.57	.99	353	1.82*	.47

Note. N = number of participants who completed the explicit measure. For Studies 1–3, means are an average between two items after standardizing each measure ($SD = 1$) while retaining a rational zero point indicating no preference. For Study 4, means represent scores on the Subtle-Blatant Prejudice Scale (Pettigrew & Meertens, 1995). More positive scores indicate a greater explicit preference for White people over Black people or more prejudiced attitudes. IAT = Implicit Association Test; GNAT = go/no-go association task.
* $p < .05$.

preferences by condition. Follow-up analyses revealed that five conditions (Practicing Counterstereotypes with the IAT, Affirming Common Humanity, Shifting Group Affiliations Under Threat, Considering Racial Injustice, Evaluative Conditioning) each elicited weaker explicit racial preferences, and two conditions (Priming an Egalitarian Mindset, Faking) elicited larger explicit racial preferences relative to a control condition, but none of these effects was replicated across studies. A random-effects meta-analytic estimate computed across all experimental conditions suggest a very weak effect on reducing explicit racial preferences relative to control ($d = .04$, 95% CI [.02, .07]). Follow-up analyses of specific interventions revealed that no meta-analytic effect size was significant at conventional hypothesis-testing thresholds ($p < .05$). At best, the evidence suggests that the interventions have an effect on explicit evaluation but that this effect is weak, and there is no evidence for differential effectiveness among interventions.

Implicit Preferences for Whites Compared With Blacks on the MC-IAT (Study 3)

In Study 3, we included a second implicit measure, the MC-IAT, to diversify measurement and to examine whether intervention effectiveness was particular to comparisons of Whites and Blacks or generalized to other racial groups (addressed in the next section). To examine the effects of interventions on implicit preferences for Whites over Blacks on the MC-IAT, we modeled implicit preferences for Whites over Blacks as a function of condition. A one-way analysis of variance revealed a significant effect of condition on implicit preferences for Whites over Blacks, $F(1933) = 3.24$, $p = .00012$, $\eta^2 = .019$. See Tables 4 and 5 for a summary of MC-IAT results. Follow-up analyses revealed that two conditions elicited significantly weaker implicit preferences for Whites over Blacks than the control condition: Faking the IAT, $t_{sattherwaite}(291.98) = 4.22$, $p = .000033$, $d = .47$, and Priming Multiculturalism, $t(307) = 2.09$, $p = .038$, $d = .25$. Nine of the 10 remaining interventions elicited weaker implicit preferences for Whites compared with Blacks but did not achieve statistical significance.

In Study 3, two interventions significantly reduced implicit preferences for Whites over Blacks on the MC-IAT, and nine interventions reduced implicit preferences for Whites over Blacks on the IAT. This different pattern of results could reflect important differences between the associations measured by the IAT and MC-IAT. The degree of racial preference elicited in the control conditions suggest that the IAT ($M = .50$, $SD = .43$), $t(181) = 15.98$, $p = 2 \times 10^{-36}$, $d = 1.19$, elicits racial preferences that are approximately double the magnitude observed on the MC-IAT ($M = .26$, $SD = .52$), $t(161) = 6.28$, $p = 2.96 \times 10^{-9}$, $d = .49$. The likely explanation for the difference in the magnitude is that

Table 4
Implicit Racial Preferences on the MC-IAT (Comparisons With White People)

Condition	White compared with Black			White compared with Hispanic			White compared with Asian		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Control	162	.26	.52	165	.28	.46	162	.08	.48
Exposure to counterstereotypical exemplars									
Vivid Counterstereotypic Scenario	164	.17	.51	166	.27	.49	164	.07	.51
Practicing an IAT With Counterstereotypical Exemplars	154	.20	.58	154	.25	.52	152	.14	.46
Shifting Group Boundaries Through Competition	149	.23	.53	148	.25	.50	147	.06	.50
Shifting Group Affiliations Under Threat	162	.23	.57	163	.26	.49	162	.15	.53
Appeals to egalitarian values									
Priming Feelings of Nonobjectivity	147	.19	.55	148	.23	.59	148	.07	.54
Instilling a Sense of Common Humanity	134	.29	.55	134	.24	.51	134	.10	.54
Priming an Egalitarian Mindset	135	.23	.52	134	.33	.43	134	.10	.48
Priming Multiculturalism	169	.13*	.52	169	.26	.52	168	.04	.54
Evaluative conditioning									
Evaluative Conditioning	128	.17	.54	129	.32	.52	129	.12	.50
Evaluative Conditioning With the GNAT	128	.20	.50	128	.16*	.56	128	.13	.49
Intentional strategies to overcome biases									
Using Implementation Intentions	152	.15	.53	152	.23	.57	152	−.02	.53
Faking the IAT	162	−.04***	.72	161	.02***	.69	162	−.22***	.65

Note. *N* = number of completed Multi-Category Implict Association Tests (MC-IATs) for the condition. MC-IAT means are *D* scores (Nosek, Bar-Anan, Sriram, & Greenwald, 2012), and positive values indicate greater preference for White people compared with the other racial ethnic group. IAT = Implicit Association Test; GNAT = go/no-go association task.
* $p < .05$. *** $p < .001$.

the MC-IAT focuses on assessment of positive associations (Nosek et al., 2014; Sriram & Greenwald, 2009), whereas the IAT is attentive to both positive and negative associations. Using a similar procedure, Nosek and colleagues (2012) found that Bad was more strongly associated with Blacks than Whites and that this association was relatively independent of associations with Good. So, with only the "good" half of the effect influencing MC-IAT performance, the effect in the control condition was just half the effect size compared with the IAT. As a consequence, an equally effective intervention would be less likely to demonstrate statisti-

Table 5
Implicit Racial Preferences on the MC-IAT (Comparisons With Non-White People)

Condition	Asian compared with Hispanic			Asian compared with Black			Hispanic compared with Black		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Control	162	.20	.50	165	.17	.50	163	.07	.51
Exposure to counterstereotypical exemplars									
Vivid Counterstereotypic Scenario	165	.24	.47	166	.08	.55	166	−.02	.51
Practicing an IAT With Counterstereotypical Exemplars	154	.21	.49	153	.11	.50	151	.01	.53
Shifting Group Boundaries Through Competition	149	.19	.47	148	.04**	.48	149	.00	.46
Shifting Group Affiliations Under Threat	163	.26	.49	162	.08	.50	162	.04	.51
Appeals to egalitarian values									
Priming Feelings of Nonobjectivity	148	.17	.48	147	.10	.48	148	−.01	.52
Instilling a Sense of Common Humanity	134	.23	.46	133	.14	.53	134	−.04*	.45
Priming an Egalitarian Mindset	135	.19	.46	134	.20	.44	135	.03	.49
Priming Multiculturalism	168	.24	.49	169	.10	.50	169	−.01	.47
Evaluative conditioning									
Evaluative Conditioning	129	.20	.47	128	.05	.50	128	.03	.49
Evaluative Conditioning With the GNAT	128	.17	.44	128	.13	.43	127	.03	.44
Intentional strategies to overcome biases	152	.30	.49	153	.08	.53	153	−.07*	.50
Using Implementation Intentions	152	.30	.49	153	.08	.53	153	−.07*	.50
Faking the IAT	162	.25	.50	162	−.01**	.52	162	−.10**	.51

Note. *N* = number of completed Multi-Category Implicit Association Tests (MC-IATs) for the condition. MC-IAT means are *D* scores (Nosek, Bar-Anan, Sriram, & Greenwald, 2012), and positive values indicate greater preference for the first named group compared with the second named group (e.g., higher scores on the Asian compared with Hispanic column reflects greater preference for Asians compared with Hispanics). IAT = Implicit Association Test; GNAT = go/no-go association task.
* $p < .05$. ** $p < .01$.

cally significant reductions toward an effect size of zero. Indeed, considering the change in terms of effect magnitude, MC-IAT effects were still reduced in intervention conditions by 31.1% of the control condition, only slightly smaller than the 35.5% reduction in IAT effects across interventions. This suggests that the weaker results on significance testing for the MC-IAT were a function of having less power to detect the smaller effects rather than the interventions being less effective. Further, associations of racial groups with both Good and Bad may be shifting as a function of these interventions, perhaps differentially so. A future investigation could examine this directly by assessing associations with both Good and Bad using the MC-IAT.

Implicit Preferences for Other Racial/Ethnic Groups (Study 3)

Do interventions that reduce the expression of implicit racial preferences for Whites over Blacks extend to other racial/ethnic outgroups? One possibility is that interventions are "localized" and shift associations about Blacks and Whites exclusively. Another possibility is that interventions shift motivations or activate egalitarian associations more generally, thus shifting evaluations toward any racial outgroup. For example, some evidence shows that prejudice reduction from intergroup contact is generalized from the primary outgroup to unrelated secondary outgroups (Pettigrew, 1997, 2009; Schmid, Hewstone, Küpper, Zick, & Wagner, 2012; Tausch et al., 2010).

To examine the effects of interventions on implicit preferences for other racial groups, we modeled implicit preferences for Whites over Asians and Whites over Hispanics on the MC-IAT as a function of condition. One-way analyses of variance revealed significant effects of condition on implicit preferences for Whites compared with Asians, $F(1929) = 5.21$, $p = 1.10 \times 10^{-8}$, $\eta^2 = .031$, and Whites compared with Hispanics, $F(1938) = 3.40$, $p = .000061$, $\eta^2 = .020$. Follow-up analyses of individual conditions indicated that faking instructions decreased implicit preferences for Whites over Asians, $t_{sattherwaite}(291.98) = 4.22$, $p < .001$, $d = .47$. Additionally, faking instructions, $t_{sattherwaite}(277.71) = 4.08$, $p < .001$, $d = .46$, and evaluative conditioning with the GNAT, $t_{sattherwaite}(234.42) = 2.03$, $p = .043$, $d = .25$, decreased implicit preferences for Whites over Hispanics. Like changes in explicit racial preference for Whites and Blacks, this evidence suggests that the interventions affected preferences between Whites and other racial groups to some degree, but only weakly so. Compared with control, interventions reduced MC-IAT effects for Whites compared with Asians by 22.9% and reduced MC-IAT effects for Whites compared with Hispanics by 16.1%. These reductions were smaller than reductions in MC-IAT effects for Whites compared with Blacks (35.5%).

We also examined the effect of condition on feeling thermometer ratings of Whites, Blacks, Asians, and Hispanics. One-way analyses of variance revealed no significant effects of condition on feeling thermometer ratings of any of the four racial/ethnic groups examined (ps $> .05$).

Pretest-Related Analyses (Study 4)

In Study 4, we used a Solomon "four group" design to test whether differential attrition could account for any of the observed effects. First, among those who completed the pretest, there was no difference in baseline implicit racial preferences between those who completed the study ($M = .39$, $SD = .48$) and those who did not ($M = .39$, $SD = .47$), $t(2991) = .11$, $p = .91$, $d = .02$, nor was there was an interaction of completed/did not with experimental condition in predicting baseline implicit preferences, $F(12, 2963) = .85$, $p = .60$, $\eta^2 = .003$. Second, considering the whole sample, there was no main effect of whether the person completed the pretest ($M = .30$, $SD = .47$) or not ($M = .29$, $SD = .52$) on the posttest IAT, $t_{sattherwaite}(4871.69) = .68$, $p = .50$, $d = .02$. However, there was a small interaction of pretest/no-pretest with condition, $F(12, 4995) = 2.57$, $p = .0021$, $\eta^2 = .006$. Participants who completed evaluative conditioning (Intervention 14) after a taking a pretest had higher posttest IAT scores than participants who completed evaluative conditioning without a pretest, $t_{sattherwaite}(349.24) = -3.13$, $p = .0019$, $d = -.32$. A similar, nonsignificant trend was found for participants who completed evaluative conditioning with the GNAT (Intervention 15), $t(390) = -1.66$, $p = .097$, $d = -.17$. Evaluative conditioning tasks may have shown reduced effectiveness after a pretest because categorizing White and Black faces on the pretest IAT decreased the relation between White faces + Bad and Black faces + Good established by those tasks. In addition, participants who were primed with feelings of nonobjectivity (Intervention 9) and completed a pretest had lower posttest IAT scores than participants who were primed with feelings of nonobjectivity without a pretest, $t(377) = 2.51$, $p = .013$, $d = .26$. This may be because reading about nonconscious biases may make more sense to participants after attempting to overcome their biases on the IAT. Third, effects across conditions were very similar whether the pretest was included as a covariate or not, with the exceptions of evaluative conditioning and priming feelings of nonobjectivity. Overall, however, pretest inclusion did not alter the substantive conclusion of any effect. In sum, we found no evidence that differential attrition could account for the observed effects, and slight evidence that including a pretest changes the effectiveness of several interventions. All pretest-related analyses conducted are available in an online supplement at https://osf.io/lw9e8/.

General Discussion

In four studies with 17,021 total participants, we investigated the comparative effectiveness of 18 interventions to reduce implicit racial preferences. All interventions are presented in Figure 1 along with their meta-analytic confidence intervals. Eight of the 17 interventions plus the faking condition were successful in reducing implicit preferences at least once, and all nine of these had 95% confidence intervals that did not include zero after meta-analytically aggregating across studies. The 18 experimental conditions, from most effective to least effective (by meta-analytic effect size) were as follows: Shifting Group Boundaries Through Competition (Intervention 6), Vivid Counterstereotypic Scenario (Intervention 4), Practicing an IAT With Counterstereotypical Exemplars (Intervention 5), Priming Multiculturalism (Intervention 13), Evaluative Conditioning With the GNAT (Intervention 15), Faking the IAT (Intervention 18), Shifting Group Affiliations Under Threat (Intervention 7), Using Implementation Intentions (Intervention 17), Evaluative Conditioning (Intervention 14), Inducing Moral Elevation (Intervention 16), Considering Racial In-

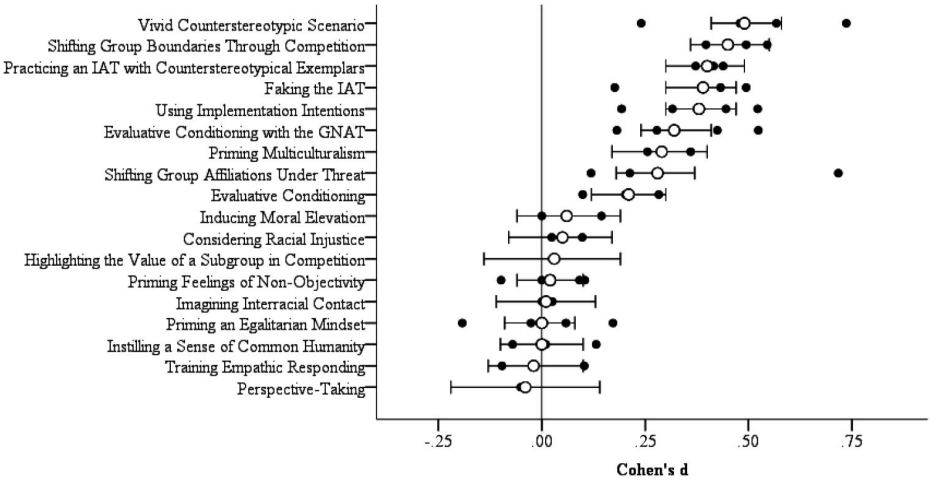

Figure 1. Effectiveness of interventions on implicit racial preferences, organized from most effective to least effective. Cohen's d = reduction in implicit preferences relative to control; White circles = the meta-analytic mean effect size; Black circles = individual study effect sizes; Lines = 95% confidence intervals around meta-analytic mean effect sizes. IAT = Implict Association Test; GNAT = go/no-go association task.

justice (Intervention 10), Highlighting the Value of a Subgroup in Competition (Intervention 8), Imagining Interracial Contact (Intervention 3), Instilling a Sense of Common Humanity (Intervention 11), Training Empathic Responding (Intervention 1), Priming Feelings of Nonobjectivity (Intervention 9), Perspective Taking (Intervention 2), and Priming an Egalitarian Mindset (Intervention 12). Using null hypothesis significance testing, the first nine conditions listed above were effective at reducing implicit preferences, whereas the last nine were ineffective. There was considerable variability in effectiveness among the successful ones (ds ranging from .16 to .51, average d = .36, SD = .10), whereas ineffective interventions were fairly homogeneous (ds ranging from −.04 to .06, average d = .0022, SD = .04).

We organized the 18 conditions into six descriptive categories: exposure to counterstereotypical exemplars, intentional strategies to overcome bias, evaluative conditioning, engaging with others' perspectives, appeals to egalitarian values, and inducing emotion. Interventions from the first three categories were especially effective at reducing implicit preferences: exposure to counterstereotypical exemplars, d = .38, 95% CI [.32, .44]; intentional strategies to overcome bias, d = .32, 95% CI [.25, .39]; and evaluative conditioning, d = .26, 95% CI [.18, .34]. Four out of five interventions that exposed participants to counterstereotypical exemplars and all of the interventions that used intentional strategies or evaluative conditioning reduced implicit preferences at least once. Interventions from the other three categories tended to be ineffective: perspective taking, d = −.01, 95% CI [−.07, .05]; appeals to egalitarian values, d = .03, 95% CI [−.04, .09]; and emotion induction, d = .06, 95% CI [−.06, .19]. None of the three interventions that led participants to engage with others' perspectives and only one of the five interventions that appealed to egalitarian values reduced implicit racial preferences, a total of one out of 11 tests of malleability showing a significant effect. This is particularly notable considering that both engagement with others' perspectives and appeals to egalitarian values have well-developed theoretical rationales and even some published evidence supporting them (Legault et al., 2011; Todd, Bodenhausen, Richeson, & Galinsky, 2011). Emotion induction was not effective either; however, only one type of emotion was induced (elevation), so it would be premature to conclude that such manipulations are not effective. Other research using emotional states to alter implicit preferences have revealed that anger and disgust can increase implicit preferences for ingroups compared with outgroups (Dasgupta et al., 2009; DeSteno et al., 2004), although no published result demonstrates an emotional state that decreases preferences for ingroups compared with outgroups.

It is important to note that these categorizations are descriptive rather than denotative of the relevant psychological mechanisms. The most appropriate interpretation of intervention effectiveness is at the level of the interventions themselves. These descriptive categories offer an opportunity to theorize about the mechanisms that may be operating effectively or ineffectively across intervention designs. As described earlier, each of the interventions has a strong rationale grounded in existing theory, and many of the interventions are adaptations of ones that have published evidence suggesting their effectiveness and identifying possible mechanisms of effectiveness. Follow-up research will clarify whether those mechanisms are related to the descriptive categories or are based on other features of the paradigms.

Insights About Effectiveness and Ineffectiveness

The current results are revealing about interventions that were ineffective as well as the interventions that were effective. In some research applications, null effects are difficult to interpret because they could be due to low power, sample characteristics, procedural details other than the intervention itself, and even the implementation of the dependent variable. The null effects in the current research cannot be accounted for by these factors. Most interventions were tested multiple times, and the large samples afforded 80% power (on average) in each study to detect effects of d = .25. Lack of effectiveness also does not mean that the theoretical

mechanism could never work. It may be that the mechanism for change would be effective with a longer intervention session, other samples, other settings, other implicit measures as dependent variables, or with a "control" condition that more closely matches the intervention rather than our "pure baseline" control condition. This would mean that there are constraints for when such interventions will be effective that are not yet clear. Finally, because published studies are biased against null results (Fanelli, 2010, 2012; Sterling, 1959; Sterling, Rosenbaum, & Weinkam, 1995), it is possible that the current literature overestimates the effectiveness of some types of interventions. Whatever the explanation, the current results raise new questions about the conditions under which the ineffective interventions could be effective. In this way, the highly powered negative results based on paradigms with strong conceptual justifications can be generative for explaining when interventions can be effective.

Notable Features of Successful Interventions

Beyond the six descriptive categories, several notable features were related to intervention success. First, 21 out of 27 successful attempts at reducing implicit preferences linked positivity with Black people *and* negativity with White people. In contrast, only six out of 27 unsuccessful attempts linked positivity with Black people and negativity with White people. This pattern is best illustrated with Shifting Group Affiliations Under Threat (Intervention 7). In Study 2, participants read a story where they belonged to a group of postwar survivors who were all Black. In Studies 3 and 4, the intervention was revised to also include profiles of a villainous rival band of survivors, who happened to be all White. Whereas the intervention failed to significantly reduce preferences in Study 2, the revised form was significantly effective in Studies 3 and 4. This bolsters evidence from Joy-Gaba and Nosek (2010) suggesting that focusing on positivity toward the devalued group in isolation may be less effective than directly contrasting between social groups. Of course, interventions that reduce relative preferences by increasing negativity toward the more positively valued group may not be desirable for application. Whether effective intervention strategies for reducing bias *should* be used in practice are ethical considerations, not scientific ones.

Prior research on explicit attitudes has revealed that involvement (the degree to which an attitude is linked with the self) can affect how persuasive messages are processed (Johnson & Eagly, 1989). Similarly, we found that high involvement was a common feature among interventions that were successful at reducing implicit preferences. For example, participants in the Vivid Counterstereotypic Scenario (Intervention 4) imagined being assaulted by a White man and rescued by a Black man. This intervention was effective in Study 1 (d = .24), but was more effective in Studies 2, 3, and 4 (ds = .48, .75, .57) after being revised to be more involving. Process data from Marini, Rubichi, and Sartori (2012) revealed that this intervention reduces implicit preferences when participants imagine they are assaulted, but not when they imagine someone else being assaulted. Another way in which interventions invoked involvement was by emphasizing competition between groups and making the relevant ingroup mostly Black. Whereas the interventions that made the participant an active member were very effective (Shifting Group Boundaries Through Competition [Intervention 6], Shifting Group Affiliations Under Threat [Intervention 7]), an intervention that only reminded participants of an ingroup association without a vivid or personally relevant narrative (Highlighting the Value of a Subgroup in Competition [Intervention 8]) was not.

An important feature of these studies was the direct comparison of interventions designed to alter the activation or expression of implicit preferences, compared with "faking" that aims to manipulate task performance itself. In contrast to self-report, prior research suggests that participants do not spontaneously fake their implicit responses (Banse et al., 2001; Kim, 2003; Steffens, 2004) without precise instructions on how to fake (Fiedler & Bluemke, 2005). In our studies, we used this prior knowledge to develop a faking instruction that was successful but was not among the most effective manipulations. Faking ranked fifth out of the nine effective manipulations. Faking presumably has few implications for "actual" change that could also lead to a change in behavior, but it has substantial value in the current studies for differentiating "bogus" interventions from others. The standard deviations in the faking condition were considerably larger than other conditions by 27% in Study 1, 44% in Study 2, 36% in Study 3, and 40% in Study 4, making it easy to detect compared with "real" interventions that presumably altered the activation or expression of implicit preferences (see Table 2). In this context, actual interventions shifting activation and expression of implicit preferences could outperform task manipulation through faking, and the latter is detectable (as "not real") through increases in variability (Cvencek et al., 2010; Röhner, Schröder-Abé, & Schütz, 2013).

The Most Effective Interventions Leverage Multiple Mechanisms

The three most effective interventions appear to leverage multiple mechanisms to increase their impact on implicit preferences (Shifting Group Boundaries Through Competition [Intervention 6], Vivid Counterstereotypic Scenario [Intervention 4], and Practicing an IAT With Counterstereotypical Exemplars [Intervention 5]). The most effective intervention, Vivid Counterstereotypic Scenario, involved the participant as the subject of the story, had the participant imagine his- or herself under a highly threatening life-or-death situation, exposed participants to counterstereotypical exemplars (malevolent White villain, dashing Black hero), and provided strategies to overcome bias (goal intentions to associate good with Black and bad with White) to reduce implicit preferences.

When change in an outcome variable is the primary focus of research, multiple mechanisms can work multiplicatively to produce powerful effects that would not arise when the primary interest is testing mechanisms in isolation. The mechanism-focused and contest approaches are a very powerful combination for translating basic research to application. In the standard experimental context, it is pragmatically difficult to investigate the additive and interactive effects of many mechanisms at once. The contest approach offers an efficient means of identifying interventions that are particularly effective, regardless of how mechanisms are combined. Knowing what works can then feed back into basic research to unpack how it works. This way, mechanism research can focus on the particular combinations that are effective rather than inefficiently examining all possible combinations with maximal experimental control.

General Ineffectiveness of Changing Explicit Racial Preferences

Interventions were ineffective at reducing explicit preferences despite the use of approaches that had been validated in explicit prejudice reduction research. One explanation is that self-reports of racial attitudes are less amenable to change due to cultural shifts in the acceptability of racial prejudice that have suppressed the self-reporting of racial prejudice (Sniderman & Piazza, 1993). Indeed, most research investigating explicit prejudice reduction toward Black people was published before 1990. However, this explanation does not account easily for the current findings. Participants were willing to express moderately strong explicit preferences for Whites compared with Blacks (*d*s $>$ .50), leaving substantial opportunity to reduce explicit prejudice. Another explanation is that the design goal of reducing implicit preferences incentivizes intervention designs that target implicit cognitions specifically. A similar research contest targeting explicit racial preferences could yield different approaches and innovation in effective interventions for explicit prejudice reduction. Lastly, explicit preferences were always assessed after implicit measurement. Although there were only a few minutes between intervention and assessment, we cannot rule out the possibility that the intervention effectiveness dissipated prior to explicit measurement.

Evidence for Shifting Racial Biases Toward Multiple Groups

Interventions that reduce implicit preferences for Whites over Blacks could operate by changing associations related to egalitarianism, or by changing associations about Blacks and Whites exclusively. In Study 3, we tested the possibility that interventions could also induce attitude change toward other racial outgroups. We found that implicit, but not explicit, preferences for Whites compared with Asians and Hispanics were reduced compared with control. Malleability in these preferences may reflect shifts in attitudes toward Asians or Hispanics, shifts in attitudes toward Whites, or both. The current design cannot distinguish among these possibilities.

Limitations

The current research allows for comparative inferences about interventions in a specific experimental context. Changes in procedure, sample, or setting could alter the overall effectiveness and relative ranking of these interventions for reducing implicit racial preferences. For example, a longer intervention format could increase the effectiveness of evaluative conditioning (Bar-Anan, De Houwer, & Nosek, 2010), but might have no effect on some other interventions. Furthermore, although we examined a wide variety of interventions, they surely are not comprehensive of all plausible interventions.

Another possibility is that interventions will vary in effectiveness with use of different implicit measures. We used two implicit measures that, despite their shared reliance on categorization, have unique psychometric properties (Nosek et al., 2014; Sriram & Greenwald, 2009). We also considered using priming-based implicit measures that may capture distinct aspects of implicit racial attitudes but did not use them in this research design. Evaluative priming (Fazio, Sanbonmatsu, Powell, & Kardes, 1986) is useful and used widely, but suffers from low reliability and very weak effect sizes compared with the IAT (e.g., $d = .07$ vs. $d = .75$ in a comparative investigation; Bar-Anan & Nosek, in press). The power required to detect change reliably with such weak effect sizes in our paradigm far exceeded available resources. The affect misattribution procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005) is more reliable than evaluative priming. However, the AMP's psychometric qualities appear to be dependent on a subset of respondents, making it less attractive as a measure of intervention effectiveness (Bar-Anan & Nosek, 2012, in press). Another issue with the implicit measures used in this design is their relative nature—changes in IAT or MC-IAT scores may reflect changes in attitudinal responses to Blacks, Whites, or both. Future investigations using implicit measures that are more adept at targeting associations for a single object (e.g., the Single-Category IAT; Karpinski & Steinman, 2006) may shed more light on the mechanisms underlying the malleability effects found in these studies. As implicit measures mature, replicating this design with new methods will increase confidence in the theoretical interpretation of what these interventions are changing.

The current research also allowed us to draw inferences about which interventions are effective within a single session but provided little information about their durability. Although most research on shifting implicit preferences is conducted with the independent and dependent variable occurring in the same experimental session, this is a general limitation for this research area (Lai, Hoffman & Nosek, 2013). Understanding the time course of intervention effectiveness has important implications for application; interventions that induce temporary change may be useful for immediate application in specific social contexts, and interventions that instill long-term change may aid in reducing discrimination across many contexts. Nonetheless, save for the elusive sleeper effect (Gillig & Greenwald, 1974; Pratkanis, Greenwald, Leippe, & Baumgardner, 1988), a prerequisite for long-term change is short-term change. As such, it is efficient to first identify effects that work in the short term so as to focus long-term investigations on interventions that show initial promise. The current work offers a basis for selection for investigating durability of effective interventions.

Conclusion

There is a demand for interventions that are effective in reducing prejudice but a paucity of applied evidence to guide practitioners toward best practices (Paluck & Green, 2009). Our research contest provides a starting point for comparative evaluation of interventions grounded in existing psychological theory and evidence. By experimentally comparing 17 interventions and a faking comparison condition for reducing implicit racial preferences, we found that interventions featuring exposure to counterstereotypical exemplars, intentional strategies to overcome biases, and evaluative conditioning were consistently more effective than ones that featured engagement with others' perspectives, appeals to egalitarian values, and elevation induction for reducing implicit preferences for Whites compared with Blacks.

Coupled with basic research clarifying the mechanisms of change, these results provide a next step toward understanding the malleability of implicit racial preferences and developing applications of that basic knowledge. Necessary steps following the current one include evaluating the durability of these intervention effects, investigating whether the change in implicit preference predicts a subsequent change in behavior, and investigating whether the effective interven-

tions can scale up to practical applications (Lai et al., in press). These steps toward application may require more intensive interventions that are longer, use multiple administrations, or use a combination of multiple strategies. The cumulative knowledge from these steps will constitute a bridge between basic research on principles for attitude change to practical application of those principles for effecting social change.

References

Algoe, S. B., & Haidt, J. (2009). Witnessing excellence in action: The "other-praising" emotions of elevation, gratitude, and admiration. *Journal of Positive Psychology, 4,* 105–127. doi:10.1080/17439760802650519

Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.

Ames, D. L., Jenkins, A. C., Banaji, A. C., & Mitchell, J. P. (2008). Taking another person's perspective increases self-referential neural processing. *Psychological Science, 19,* 642–644. doi:10.1111/j.1467-9280.2008.02135.x

Avenanti, A., Sirigu, A., & Aglioti, S. M. (2010). Racial bias reduces empathic sensorimotor resonance with other-race pain. *Current Biology, 20,* 1018–1022. doi:10.1016/j.cub.2010.03.071

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes toward homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie, 48,* 145–160.

Bar-Anan, Y., De Houwer, J., & Nosek, B. A. (2010). Evaluative conditioning and conscious knowledge of contingencies: A correlational investigation with large samples. *Quarterly Journal of Experimental Psychology, 63,* 2313–2335. doi:10.1080/17470211003802442

Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the affective misattribution procedure. *Personality and Social Psychology Bulletin, 38,* 1194–1208. doi:10.1177/0146167212446835

Bar-Anan, Y., & Nosek, B. A. (in press). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*.

Bargh, A. J. (1999). The unbearable automaticity of being. *American Psychologist, 54,* 462–479. doi:10.1037/0003-066X.54.7.462

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6,* 242–261. doi:10.1207/S15327957PSPR0603_8

Bonilla-Silva, E., Lewis, A., & Embrick, D. G. (2004). "I did not get that job because of a black man. . .": The story lines and testimonies of color-blind racism. *Sociological Forum, 19,* 555–581. doi:10.1007/s11206-004-0696-3

Brandstätter, V., Lengfelder, A., & Gollwitzer, P. M. (2001). Implementation intentions and efficient action initiation. *Journal of Personality and Social Psychology, 81,* 946–960. doi:10.1037/0022-3514.81.5.946

Crisp, R. J., & Turner, R. N. (2009). Can imagined interactions produce positive perceptions? Reducing prejudice through simulated social contact. *American Psychologist, 64,* 231–240. doi:10.1037/a0014718

Cvencek, D., Greenwald, A. G., Brown, A., Gray, N., & Snowden, R. (2010). Faking of the Implicit Association Test is statistically detectable and partly correctable. *Basic and Applied Social Psychology, 32,* 302–314. doi:10.1080/01973533.2010.519236

Dadds, M. R., Bovbjerg, D. H., Redd, W. H., & Cutmore, R. H. (1997). Imagery in human classical conditioning. *Psychological Bulletin, 122,* 89–103. doi:10.1037/0033-2909.122.1.89

Dasgupta, N. (2009). Mechanisms underlying malleability of implicit prejudice and stereotypes: The role of automaticity versus cognitive control. In T. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 267–285). Mahwah, NJ: Erlbaum.

Dasgupta, N., DeSteno, D., Williams, L. A., & Hunsinger, M. (2009). Fanning the flames of prejudice: The influence of specific incidental emotions on implicit prejudice. *Emotion, 9,* 585–591. doi:10.1037/a0015961

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81,* 800–814. doi:10.1037/0022-3514.81.5.800

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135,* 347–368. doi:10.1037/a0014211

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127,* 853–869. doi:10.1037/0033-2909.127.6.853

DeSteno, D., Dasgupta, N., Bartlett, M. Y., & Cajdric, A. (2004). Prejudice from thin air: The effect of emotion on automatic intergroup attitudes. *Psychological Science, 15,* 319–324. doi:10.1111/j.0956-7976.2004.00676.x

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18. doi:10.1037/0022-3514.56.1.5

Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology, 60,* 817–830. doi:10.1037/0022-3514.60.6.817

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33,* 510–540. doi:10.1006/jesp.1997.1331

Eibach, R. P., & Ehrlinger, J. (2006). "Keep your eyes on the prize": Reference points and racial differences in assessing progress toward equality. *Personality and Social Psychology Bulletin, 32,* 66–77. doi:10.1177/0146167205279585

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS ONE, 5,* e10271. doi:10.1371/journal.pone.0010271

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90,* 891–904. doi:10.1007/s11192-011-0494-7

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69,* 1013–1027. doi:10.1037/0022-3514.69.6.1013

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50,* 229–238. doi:10.1037/0022-3514.50.2.229

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology, 27,* 307–316. doi:10.1207/s15324834basp2704_3

Finlay, K. A., & Stephan, W. G. (2000). Improving intergroup relations: The effects of empathy on racial attitudes. *Journal of Applied Social Psychology, 30,* 1720–1737. doi:10.1111/j.1559-1816.2000.tb02464.x

Foroni, F., & Mayr, U. (2005). The power of a story: New, automatic, associations from a single reading of a short scenario. *Psychonomic Bulletin & Review, 12,* 139–144. doi:10.3758/BF03196359

Gaertner, S. L., & Dovidio, J. F. (2000). *Reducing intergroup bias: The common ingroup identity model*. Philadelphia, PA: Psychology Press. doi:10.4135/9781446218617.n9

Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., & Rust, M. C. (1993). The common ingroup identity model: Recategorization and the reduction of intergroup bias. *European Review of Social Psychology, 4,* 1–26. doi:10.1080/14792779343000004

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692–731. doi:10.1037/0033-2909.132.5.692

Gawronski, B., & Sritharan, R. (2010). Formation, change, and contextualization of mental associations: Determinants and principles of variations in implicit measures. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 216–240). New York, NY: Guilford Press.

Gillig, P. M., & Greenwald, A. G. (1974). Is it time to lay the "sleeper effect" to rest? *Journal of Personality and Social Psychology, 29,* 132–139. doi:10.1037/h0035744

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54,* 493–503. doi:10.1037/0003-066X.54.7.493

Gollwitzer, P. M., & Schaal, B. (1998). Metacognition in action: The importance of implementation intentions. *Personality and Social Psychology Review, 2,* 124–136. doi:10.1207/s15327957pspr0202_5

Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for Black and White patients. *Journal of General Internal Medicine, 22,* 1231–1238. doi:10.1007/s11606-007-0258-5

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4–27. doi:10.1037/0033-295X.102.1.4

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74,* 1464–1480. doi:10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test I: An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197–216. doi:10.1037/0022-3514.85.2.197

Haidt, J. (2003). Elevation and the positive psychology of morality. In C. L. M. Keyes & J. Haidt (Eds.), *Flourishing: Positive psychology and the life well-lived* (pp. 275–289). Washington, DC: American Psychological Association. doi:10.1037/10594-012

Johnson, B. T., & Eagly, A. H. (1989). Effects of involvement on persuasion: A meta-analysis. *Psychological Bulletin, 106,* 290–314. doi:10.1037/0033-2909.106.2.290

Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology, 33,* 1–27. doi:10.1111/j.2044-8309.1994.tb01008.x

Joy-Gaba, J. A., & Nosek, B. A. (2010). The surprisingly limited malleability of implicit racial evaluations. *Social Psychology, 41,* 137–146. doi:10.1027/1864-9335/a000020

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91,* 16–32. doi:10.1037/0022-3514.91.1.16

Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming status of dual cognitive structures. *Journal of Personality and Social Psychology, 55,* 893–905. doi:10.1037/0022-3514.55.6.893

Kim, D. (2003). Voluntary controllability of the Implicit Association Test *Social Psychology Quarterly, 66,* 83–96. doi:10.2307/3090143

Lai, C. K., Haidt, J., & Nosek, B. A. (2013a). Moral elevation reduces prejudice against gay men. *Cognition and Emotion*. Advance online publication. doi:10.1080/02699931.2013.861342

Lai, C. K., Hoffman, K. M., & Nosek, B. A. (2013). Reducing implicit prejudice. *Social and Personality Psychology Compass, 7,* 315–330. doi:10.1111/spc3.12023

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. L., Teachman, B. A., . . . Nosek, B. A. (in press). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*.

Legault, L., Gutsell, J. N., & Inzlicht, M. (2011). Ironic effects of anti-prejudice messages: How motivational interventions can reduce (but also increase) prejudice. *Psychological Science, 22,* 1472–1477. doi:10.1177/0956797611427918

Mann, N. H., & Kawakami, K. (2012). The long, steep path to equality: Progressing on egalitarian goals. *Journal of Experimental Psychology: General, 141,* 187–197. doi:10.1037/a0025602

Marini, M., Rubichi, S., & Sartori, G. (2012). The role of self-involvement in shifting IAT effects. *Experimental Psychology, 28,* 348–354.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37,* 435–442. doi:10.1006/jesp.2000.1470

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience, 17,* 1306–1315. doi:10.1162/0898929055002418

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General, 132,* 455–469. doi:10.1037/0096-3445.132.3.455

Moskowitz, G. B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology, 47,* 103–116. doi:10.1016/j.jesp.2010.08.014

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19,* 625–666. doi:10.1521/soco.19.6.625.20886

Nosek, B. A., Bar-Anan, Y., Sriram, N., & Greenwald, A. G. (2012). *Understanding and using the Brief Implicit Association Test: I. Recommended scoring procedures*. Unpublished manuscript.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality Social Psychology Bulletin, 31,* 166–180. doi:10.1177/0146167204271418

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). New York, NY: Psychology Press.

Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology, 54,* 14–29. doi:10.1027/1618-3169.54.1.14

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., . . . Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18,* 36–88. doi:10.1080/10463280701489053

Nosek, B. A., Sriram, N., Smith, C. T., & Bar-Anan, Y. (2014). *The multi-category Implicit Association Test*. Unpublished manuscript.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12,* 413–417. doi:10.1111/1467-9280.00376

Olson, M. A., & Fazio, R. H. (2002). Implicit acquisition and manifestation of classically conditioned attitudes. *Social Cognition, 20,* 89–104. doi:10.1521/soco.20.2.89.20992

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically-activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin, 32,* 421–433. doi:10.1177/0146167205284004

Olsson, A., Ebert, J. P., Banaji, M. R., & Phelps, A. E. (2005). The role of social groups in the persistence of learned fear. *Science, 309,* 785–787. doi:10.1126/science.1113551

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology, 60,* 339–367. doi:10.1146/annurev.psych.60.110707.163607

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89,* 277–293. doi: 10.1037/0022-3514.89.3.277

Pettigrew, T. F. (1997). Generalized intergroup contact effects on prejudice. *Personality and Social Psychology Bulletin, 23,* 173–185. doi: 10.1177/0146167297232006

Pettigrew, T. F. (2009). Do intergroup contact effects spread to noncontacted outgroups? *Social Psychology, 40,* 55–65. doi:10.1027/1864-9335.40.2.55

Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in Western Europe. *European Journal of Social Psychology, 25,* 57–75. doi:10.1002/ejsp.2420250106

Pratkanis, A. R., Greenwald, A. G., Leippe, M. R., & Baumgardner, M. H. (1988). In search of reliable persuasion effects: III. The sleeper effect is dead: Long live the sleeper effect. *Journal of Personality and Social Psychology, 54,* 203–218. doi:10.1037/0022-3514.54.2.203

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology, 43,* 565–578. doi:10.1016/j.jesp.2006.05.011

Ranganath, K. A., & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately; explicit attitude generalization takes time. *Psychological Science, 19,* 249–254. doi:10.1111/j.1467-9280.2008.02076.x

Ratliff, K. A., & Nosek, B. A. (2011). Negativity and outgroup biases in attitude formation and transfer. *Personality and Social Psychology Bulletin, 37,* 1692–1703. doi:10.1177/0146167211420168

Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology, 40,* 417–423. doi:10.1016/j.jesp.2003.09.002

Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review, 10,* 336–353. doi:10.1207/s15327957pspr1004_4

Röhner, J., Schröder-Abé, M., & Schütz, A. (2013). What do fakers actually do to fake the IAT? An investigation of faking strategies under different faking conditions. *Journal of Research in Personality, 47,* 330–338. doi:10.1016/j.jrp.2013.02.009

Rooth, D. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics, 17,* 523–534. doi:10.1016/j.labeco.2009.04.005

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology, 81,* 856–868. doi: 10.1037/0022-3514.81.5.856

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology, 91,* 995–1008. doi:10.1037/0022-3514.91.6.995

Schmid, K., Hewstone, M., Küpper, B., Zick, A., & Wagner, U. (2012). Secondary transfer effects of intergroup contact: A cross-national comparison in Europe. *Social Psychology Quarterly, 75,* 28–51. doi: 10.1177/0190272511430235

Schwarz, N. (1998). Accessible content and accessibility experiences: The interplay of declarative and experiental information in judgment. *Personality and Social Psychology Review, 2,* 87–99. doi:10.1207/s15327957pspr0202_2

Schwarz, N., Bless, H., Fritz, S., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61,* 195–202. doi:10.1037/0022-3514.61.2.195

Sears, D. O., & Henry, P. J. (2005). Over thirty years later: A contemporary look at symbolic racism. *Advances in Experimental Social Psychology, 37,* 95–150. doi:10.1016/S0065-2601(05)37002-X

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York, NY: Cambridge University Press. doi:10.1017/CBO9781139175043

Sniderman, P. M., & Piazza, T. (1993). *The scar of race*. Cambridge, MA: Harvard University Press.

Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin, 46,* 137–150. doi:10.1037/h0062958

Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology, 56,* 283–294. doi:10.1027/1618-3169.56.4.283

Sritharan, R., & Gawronski, B. (2010). Changing implicit and explicit prejudice: Insights from the associative-propositional evaluation Model. *Social Psychology, 41,* 113–123. doi:10.1027/1864-9335/a000017

Steffens, M. C. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology, 51,* 165–179. doi:10.1027/1618-3169.51.3.165

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—Or vice versa. *Journal of the American Statistical Association, 54,* 30–34.

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician, 49,* 108–112.

Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. *Personality and Social Psychology Bulletin, 34,* 1332–1345. doi:10.1177/0146167208321269

Tajfel, H., & Turner, J. C. (1979). An intergrative theory of intergroup conflict. In W. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–48). Pacific Grove, CA: Brooks/Cole.

Tausch, N., Hewstone, M., Kenworthy, J. B., Psaltis, C., Schmid, K., Popan, J. R., . . . Hughes, J. (2010). Secondary transfer effects of intergroup contact: Alternative accounts and underlying processes. *Journal of Personality and Social Psychology, 99,* 282–302. doi: 10.1037/a0018553

Todd, A. R., Bodenhausen, G. V., Richeson, J. A., & Galinsky, A. D. (2011). Perspective taking combats automatic expressions of racial bias. *Journal of Personality and Social Psychology, 100,* 1027–1042. doi: 10.1037/a0022308

Turner, R. N., & Crisp, R. J. (2010). Imagining intergroup contact reduces implicit prejudice. *British Journal of Social Psychology, 49,* 129–142. doi:10.1348/014466609X419901

Turner, R. N., Crisp, R. J., & Lambert, E. (2007). Imagining intergroup contact can improve intergroup attitudes. *Group Processes Intergroup Relations, 10,* 427–441. doi:10.1177/1368430207081533

Uhlmann, E. L., & Cohen, G. L. (2007). "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes, 104,* 207–223. doi: 10.1016/j.obhdp.2007.07.001

(*Appendix follows*)

Appendix

Contest Inclusion Criteria

Condition	Study 1			Study 2			Study 3			Study 4		
	<=5m	Done	Valid	<=5m	Done	Valid	<=5m	Done	Valid	<=5m	Done	Valid
Control		87%	97%		85%	97%		84%	95%		91%	97%
Engaging with others' perspectives												
Training Empathic Responding	*57%*	72%	96%	*66%*	80%	98%						
Perspective Taking	94%	71%	97%									
Imagining Interracial Contact	95%	70%	96%	*84%*	76%	98%						
Exposure to counterstereotypical exemplars												
Vivid Counterstereotypic Scenario	98%	84%	96%	89%	81%	96%	92%	77%	96%	91%	85%	98%
Practicing an IAT With Counterstereotypical Exemplars	99%	81%	96%	*58%*	68%	95%	*82%*	75%	98%	86%	83%	97%
Shifting Group Boundaries Through Competition				*72%*	84%	96%	89%	77%	98%	92%	77%	95%
Shifting Group Affiliations Under Threat				86%	82%	97%	87%	80%	96%	88%	85%	98%
Highlighting the Value of a Subgroup in Competition	100%	81%	97%									
Appeals to egalitarian values												
Priming Feelings of Nonobjectivity	99%	70%	97%	99%	85%	98%	*71%*	78%	97%	72%	80%	97%
Considering Racial Injustice	*81%*	71%	*94%*	*76%*	74%	98%						
Instilling a Sense of Common Humanity				*85%*	72%	96%	91%	66%	95%	79%	74%	98%
Priming an Egalitarian Mindset	99%	80%	98%	93%	64%	96%	96%	70%	96%	98%	76%	97%
Priming Multiculturalism							*76%*	70%	98%	77%	80%	98%
Evaluative conditioning												
Evaluative Conditioning	90%	*35%*	98%	87%	59%	97%	*82%*	72%	97%	87%	76%	99%
Evaluative Conditioning With the GNAT	*6%*	64%	98%	88%	65%	97%	94%	61%	95%	94%	73%	96%
Inducing emotion												
Inducing Moral Elevation	95%	62%	98%	*77%*	65%	99%						
Intentional strategies to overcome biases												
Using Implementation Intentions	98%	80%	97%	*83%*	71%	98%	*82%*	73%	98%	86%	79%	99%
Faking the IAT	99%	81%	97%	*72%*	75%	96%	*78%*	72%	98%	91%	76%	98%

Note. <= 5m = Of participants who completed an intervention, percentage of participants who completed the intervention within 5 min; Done = percentage of participants who completed the Implicit Association Test (IAT); Valid = Of participants who completed the IAT, percentage who produced valid on-excluded data. Italicized numbers indicate a violation of a contest inclusion criterion. GNAT = go/go-no association task.

Received December 27, 2012
Revision received January 9, 2014
Accepted January 10, 2014 ■