Journal of Experimental Social Psychology 48 (2012) 1267–1278

Contents lists available at SciVerse ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Long-term reduction in implicit race bias: A prejudice habit-breaking intervention ☆

Patricia G. Devine *, Patrick S. Forscher, Anthony J. Austin [1], William T.L. Cox

Psychology Department, University of Wisconsin, Madison, WI, USA

HIGHLIGHTS

- ► We developed an intervention to produce long-term reductions in implicit race bias.
- ► The intervention produced reductions in implicit bias that lasted up to 8 weeks.
- ► The intervention also increased awareness of bias and concern about discrimination.
- ► Our results raise the hope of reducing the pernicious effects of implicit race bias.

ARTICLE INFO

Article history:
Received 28 March 2012
Revised 11 June 2012
Available online 20 July 2012

Keywords:
Prejudice
Stereotyping
Intervention
Reduction
Implicit bias
Self-regulation

ABSTRACT

We developed a multi-faceted prejudice habit-breaking intervention to produce long-term reductions in implicit race bias. The intervention is based on the premise that implicit bias is like a habit that can be broken through a combination of awareness of implicit bias, concern about the effects of that bias, and the application of strategies to reduce bias. In a 12-week longitudinal study, people who received the intervention showed dramatic reductions in implicit race bias. People who were concerned about discrimination or who reported using the strategies showed the greatest reductions. The intervention also led to increases in concern about discrimination and personal awareness of bias over the duration of the study. People in the control group showed none of the above effects. Our results raise the hope of reducing persistent and unintentional forms of discrimination that arise from implicit bias.

Introduction

Despite encouraging trends suggesting that racial prejudice in the U. S. has waned in the last half century (Gaertner & Dovidio, 1986; Schuman, Steeh, Bobo, & Krysan, 1997), widespread evidence suggests that Black people face continuing discrimination and have more adverse outcomes than White people across a variety of domains related to success and well-being (e.g., Bertrand & Mullainathan, 2004; Bradford, Newkirk, & Holden, 2009; Mitchell, Haw, Pfeifer, & Meissner, 2005; Steele, 1997; Vontress, Woodland, & Epp, 2007). The paradox of persistent racial inequalities amid improving racial attitudes has led to a search for factors underlying ongoing discrimination. Several theorists have implicated implicit race biases, which are automatically activated and often unintentional, as major contributors to the perpetuation of discrimination (e.g., Devine, 1989; Fiske, 1998; Gaertner & Dovidio, 1986).

Supporting this claim, accumulating evidence reveals that implicit biases are linked to discriminatory outcomes ranging from the seemingly mundane, such as poorer quality interactions (McConnell & Leibold, 2001), to the undeniably consequential, such as constrained employment opportunities (Bertrand & Mullainathan, 2004) and a decreased likelihood of receiving life-saving emergency medical treatments (Green et al., 2007). Many theorists argue that implicit biases persist and are powerful determinants of behavior precisely because people lack personal awareness of them and they can occur despite conscious nonprejudiced attitudes or intentions (Bargh, 1999; Devine, 1989; Gaertner & Dovidio, 1986). This process leads people to be unwittingly complicit in the perpetuation of discrimination.

The reality of lingering racial disparities, combined with the empirically established links between implicit bias and pernicious discriminatory outcomes, has led to a clarion call for strategies to reduce these biases (Fiske, 1998; Smedley, Stith, & Nelson, 2003). In response, the field has witnessed an explosion of empirical efforts to reduce implicit biases (Blair, 2002). These efforts have yielded a number of easy-to-implement strategies, such as taking the perspective of stigmatized others (Galinsky & Moskowitz, 2000) and imagining counter-stereotypic examples (Blair, Ma, & Lenton, 2001; Dasgupta & Greenwald, 2001), that lead to

☆ We thank Markus Brauer and Carlie Allison for their comments on a previous version of this paper. We also thank Becky McGill, Rachel Nitzarim, Julia Salomon, and Chelsea Wenzlaff for their help in running the experiment reported in this paper. Preparation of this article was supported by NIH grant R01 GM088477.

* Corresponding author at: Psychology Department, University of Wisconsin-Madison, 1202 W Johnson St, Madison, WI 53706, USA.
E-mail address: pgdevine@wisc.edu (P.G. Devine).
[1] Now at the University of Chicago, Chicago, IL, USA.

0022-1031/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2012.06.003

substantial reductions in implicit bias, at least for a short time (i.e., up to 24 hours). These strategies yield reductions in implicit bias even though people use the strategies at the experimenter's behest, with no intention to reduce implicit bias. It is unclear, however, whether such incidental reductions in implicit bias are enduring or whether people could intentionally implement such strategies in the service of a long-term goal to reduce implicit bias.

Although there is no direct evidence about whether one-shot strategies used at another's behest could produce enduring change, some general dual-process theories in psychology (e.g., Epstein, 1994; Smith & DeCoster, 2000; Strack & Deutsch, 2004) suggest that such reductions are likely to be highly contextual and short-lived. According to these theories, implicit and explicit processes are supported by fundamentally different psychological systems. Although the explicit system can change quickly and is relatively context-independent, the implicit system is highly contextual and only changes in an enduring way after considerable time, effort, and / or intensity of experience. Thus, because one-shot interventions must counteract a large accretion of associative learning, they are unlikely to produce enduring change in the implicit system. Such change is likely only after the application of considerable goal-directed effort over time.

The preceding analysis is consistent with Devine's habit-breaking analysis of prejudice reduction, which argues that overcoming prejudice is a protracted process that requires considerable effort in the pursuit of a nonprejudiced goal (Devine, 1989; Devine & Monteith, 1993; Devine, Monteith, Zuwerink, & Elliot, 1991; Monteith, 1993). This model likens implicit biases to deeply entrenched habits developed through socialization experiences. “Breaking the habit” of implicit bias therefore requires learning about the contexts that activate the bias and how to replace the biased responses with responses that reflect one's nonprejudiced goals.

Supporting the prejudice habit-breaking framework, considerable evidence demonstrates that, when they believe they have acted with bias, people who endorse values opposed to prejudice are motivated to inhibit the expression of implicit bias by seeking out information and putting effort into tasks they believe would help them break the prejudice habit (Amodio, Devine, & Harmon-Jones, 2007; Monteith, 1993; Plant & Devine, 2009). In addition, when these people act with prejudice, they experience guilt (Devine et al., 1991), which instigates self-regulatory efforts to disrupt automatic bias and prevent future expressions of bias (Amodio et al., 2007; Monteith, 1993). Although this evidence is consistent with the prejudice habit-breaking framework, extant research has not yet examined whether interventions can produce long-term implicit bias reductions, nor has it clearly specified the type of effort required to yield such reductions. The goal of the present work is to address these shortcomings and to develop an intervention that engages intentional effort to produce enduring reductions in implicit race bias.

Multifaceted prejudice habit-breaking intervention

Devine and colleagues (Devine & Monteith, 1993; Plant & Devine, 2009) argue that the motivation to break the prejudice habit stems from two sources. First, people must be *aware* of their biases and, second, they must be *concerned* about the consequences of their biases before they will be motivated to exert effort to eliminate them. Furthermore, people need to know when biased responses are likely to occur and how to replace those biased responses with responses more consistent with their goals.

The present work synthesizes insights from the prejudice habit model and implicit bias reduction strategies to develop an intervention to help people reduce implicit biases and “break the prejudice habit”. The multifaceted nature of the present intervention has conceptual parallels to approaches in several other areas, such as health behavior change (Prochaska & Velicer, 1997), cognitive behavior therapy (Beck & Alford, 2009; Cox, Abramson, Devine, & Hollon, 2012), and the fundamentals of adult learning (Howell, 1982; Kaufman, 2003). We tested this intervention in a three-month longitudinal study, comparing a group of people who completed the intervention to a control group who did not.

To ensure situational awareness of their bias, all participants completed a measure of implicit bias and received feedback about their level of bias. People assigned to the intervention group were also presented with a bias education and training program, the goals of which were to evoke a general concern about implicit biases and train people to eliminate such biases. The education component likened the expression of implicit biases to a habit and provided information linking implicit bias to discriminatory behaviors across a wide range of settings (e.g., interpersonal, employment, health). The training component described how to apply a variety of bias reduction strategies in daily life. Because the goal of our intervention was to engage a general self-regulatory process, we did not present the strategies in separate conditions to test each strategy's relative effectiveness. Instead, the training section presented participants with a wide array of strategies, enabling participants to flexibly choose the strategies most applicable to different situations in their lives. As part of the intervention, participants were prompted to report and reflect on their strategy use in the weeks between implicit bias assessments. We predicted that only people who received the intervention would translate their situational awareness into chronic awareness of biases in themselves and in society, thereby flipping the self-regulatory switch that motivates strategy use and reduces implicit bias.

To evaluate the effectiveness of the intervention, we examined its impact on an indicator of implicit bias and a variety of explicit measures longitudinally. We used the Black–White Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) as our measure of implicit race bias. The explicit measures included established measures of racial attitudes (Brigham, 1993), the sources of one's motivation for responding without prejudice (Plant & Devine, 1998), and whether one believes one's own behavior is more biased than appropriate (Monteith & Voils, 1998). Because 90% of our sample had a pro-White bias on the Black–White IAT, the latter served as a measure of awareness of one's tendency to respond with prejudice. In addition, because the intervention included education about the adverse effects of discrimination, we developed a measure assessing concern about discrimination in society. For both the intervention and control groups, all measures were assessed prior to the intervention manipulation and at two time points after the manipulation. We also asked the intervention group participants a variety of questions immediately after the education and training program about the strategies they had learned, and, in the weeks following the administration of the intervention, we asked them some open-ended questions about their use of the strategies.

Our design has five major strengths. First, it allows us to assess the intervention's effects on a rich array of variables (implicit and explicit) that are theoretically important to the reduction of race bias. Second, it enables us to examine whether the intervention's effects on these variables persisted or changed over time. Third, we have an opportunity to evaluate whether reported strategy use is associated with reductions in implicit bias. Fourth, in the control group, we can assess whether feedback about one's level of implicit bias leads to reductions in implicit bias without a multifaceted intervention. Finally, we can examine whether any of the explicit measurements taken at two times, prior to and after the intervention manipulation, moderate the effect of the intervention on implicit bias. A moderation effect with a measure taken prior to the intervention would suggest that the construct is related to learning processes during the intervention, while a moderation effect with a measure taken after the intervention would suggest that the construct is involved in the deployment of the bias-reducing strategies. Together, these two sets of moderation analyses can yield insight into two different aspects of the bias reduction process.

Method

Participants and design

The participants were 91 non-Black introductory psychology students (67% female, 85% White),[2] who completed a 12-week longitudinal study for course credit (see Fig. 1). Attrition rates were low, never exceeding 10% at any time point. Participants were randomly assigned to either a control condition ($n=38$) or an intervention condition ($n=53$), with more people assigned to the intervention condition to provide greater power for analyses using the strategies measures and because we anticipated greater attrition in the intervention condition (but, attrition rates did not vary across condition at any time point, all $ps>.23$). Throughout the study, participants completed the Black–White IAT and several explicit measures, described below. The IAT was administered in the lab at three time points: just prior to the intervention manipulation (baseline) and 4 and 8 weeks after the manipulation. The explicit measures were also administered at three time points: 4 weeks prior to the manipulation in a classroom setting (baseline) and 2 and 6 weeks after the manipulation via email. The intervention group participants also completed measures of their reactions to the strategies (e.g., perceived likelihood of use, perceived opportunity to use) immediately following the intervention. Finally, intervention group participants completed free-response questions about their experiences using the strategies 2 and 6 weeks after the intervention.

Implicit measure

Implicit race bias was measured with the Black–White IAT. The IAT is a dual-categorization task that has good psychometric properties (Cunningham, Preacher, & Banaji, 2001; Hofmann et al., 2005) and is linked to basic neural and affective processes relevant to implicit race bias (Cunningham, Raye, & Johnson, 2004; Phelps et al., 2000). Additionally, in intergroup contexts, the IAT is a strong predictor of discriminatory behavior and a better predictor than parallel explicit measures (Greenwald, Poehlman, Uhlmann, & Banaji, 2009; but see Blanton et al., 2009 for an opposing view).

In the Black–White IAT, people categorize sequentially presented stimuli based on whether they are pleasant or unpleasant words or Black or White faces by pressing keys on the left or right side of the keyboard. Underlying the race IAT is the assumption that, to the extent that negative valence is associated with Black people (and positive valence with White people), when Black faces and unpleasant words (and White faces and pleasant words) are paired together on the same response key (*compatible trials*), the task should be easier than with the reverse pairings (*incompatible trials*). Response times on the compatible and incompatible trials are used to compute *D*-scores (see Greenwald, Nosek, & Banaji, 2003). Higher *D*-scores indicate that participants more easily associate Black faces with unpleasant words (and White faces with pleasant words) than the reverse (baseline $M=.46$, $SD=.39$, skew $=.16$, split-half reliability $=.60$).

Explicit measures

We assessed a variety of explicit measures that have been implicated in the bias-reducing process, including racial attitudes, the sources of motivation to respond without prejudice, prejudice-relevant discrepancies, and concern about discrimination in society.

[2] We recruited non-Black participants of all races because of evidence that non-Black participants of all races show similar levels of implicit anti-Black bias (Nosek, Banaji, & Greenwald, 2002). On the basis of this evidence, we also included the non-Black participants in all analyses.

Racial attitudes

Racial attitudes were assessed using the Attitudes Towards Blacks scale (ATB; Brigham, 1993). The ATB has 20 items, each assessed on a 1 (*strongly disagree*) to 7 (*strongly agree*) scale. Responses to the items are averaged together, and higher scores indicate more positive attitudes towards Blacks and therefore, less explicit race bias (baseline $M=7.57$, $SD=.68$, skew $=-.53$, $\alpha=.82$).

Motivations to respond without prejudice

Plant and Devine (1998) distinguish between internal motivation to respond without prejudice, which is primarily driven by personal values and the belief that prejudice is wrong, and external motivation to respond without prejudice, which is primarily driven by a desire to escape social sanctions. We used the Internal Motivation Scale and External Motivation Scale (IMS and EMS; Plant & Devine, 1998) to assess these separate motivations. Both the IMS and EMS are composed of 5 items, each assessed on a 1 (*strongly disagree*) to 9 (*strongly agree*) scale. For each scale, participants' responses on the items are averaged together such that higher scores indicate more motivation to respond without prejudice (baseline IMS $M=7.03$, $SD=1.17$, skew $=-.95$, $\alpha=.72$, EMS $M=5.43$, $SD=1.78$, skew $=-.01$, $\alpha=.80$).

Shoulds, woulds, and discrepancies

The discrepancy scale measures the extent to which people predict they would act with more prejudice than they believe is appropriate. Because most people do have some implicit bias, it can also be used as an indicator of one's personal awareness of one's bias (Monteith & Voils, 1998). The full scale is composed of *should* and *would* subscales, whose items use a 1 (*strongly disagree*) to 7 (*strongly agree*) scale. The *should* subscale asks people how they believe they should act, feel, or think in response to nine interpersonal intergroup situations (e.g., I should feel uncomfortable sitting next to a Black person on a bus). Responses to the items are averaged together such that higher numbers indicate a standard that permits more prejudice (baseline $M=1.98$, $SD=.68$, skew $=1.68$, $\alpha=.62$). The *would* subscale asks people to predict how they would actually act, feel, or think in each of the situations (e.g., I would feel uncomfortable sitting next to a Black person on a bus). Responses to the items are averaged together such that higher numbers indicate predictions that the person would act with greater prejudice (baseline $M=3.18$, $SD=.95$, skew $=.10$, $\alpha=.77$). Prejudice-relevant discrepancies are calculated by first subtracting the score for each *should* from its corresponding *would*, then averaging the resulting differences. For this computed score, more positive numbers indicate a greater belief that one would act with more prejudice than one believes is appropriate in intergroup situations (baseline $M=1.20$, $SD=.86$, skew $=.17$, $\alpha=.65$).

Concern about discrimination

One of the goals of the intervention was to educate people about the existence of discrimination. We therefore developed a measure of beliefs that discrimination is a problem in society. This measure comprises four items, each measured on a 1 (*strongly disagree*) to 10 (*strongly agree*) scale. Item responses were averaged, resulting in a score for which higher numbers indicate greater concern (see Appendix A for the full measure; baseline $M=6.08$, $SD=1.22$, skew $=.15$, $\alpha=.86$).

Strategy measures

During the first lab session, following the intervention manipulation, the intervention group completed a set of Likert-type questions designed to assess the participants' reactions to each of the individual strategies. These questions used a 1 (*not at all*) to 7 (*very much*) scale and included items assessing the perceived likelihood of using each strategy, the willingness to use each strategy, the perceived difficulty of implementing each strategy, the perceived effectiveness of each strategy, and the perceived opportunities to use each strategy. The participants' responses to each item for all six techniques were averaged to obtain mean likelihood ($M=4.56$, $SD=1.09$, skew $=-.21$, $\alpha=.87$), willingness ($M=$

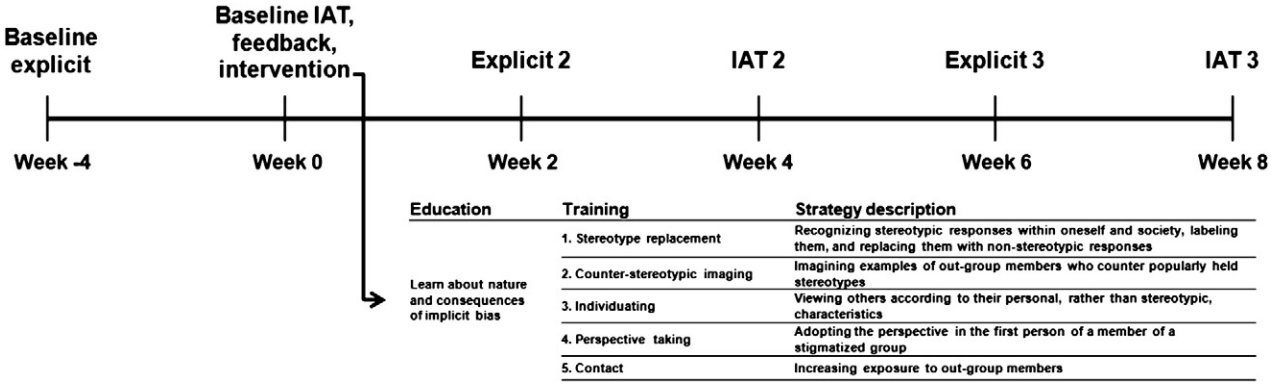

Fig. 1. Study timeline. The Black–White Implicit Association Test (IAT) was administered at 3 time points: just prior to the intervention manipulation and 4 and 8 weeks after the manipulation. The explicit measures, consisting of the Attitudes Towards Blacks (ATB) scale, the Internal and External Motivation Scales (IMS and EMS), the prejudice-relevant discrepancies scale, and the concern about discrimination scale, were also administered at three points: 4 weeks prior to the intervention manipulation during a mass survey, and 2 and 6 weeks after the manipulation.

5.78, $SD = .86$, skew $= -.20$, $\alpha = .86$), difficulty ($M = 3.74$, $SD = 1.27$, skew $= .28$, $\alpha = .85$), effectiveness ($M = 4.88$, $SD = .72$, skew .03, $\alpha = .68$), and opportunity scores ($M = 4.34$, $SD = 1.19$, skew $= -.15$, $\alpha = .83$), on which higher numbers indicate higher likelihood, willingness, perceived difficulty, perceived effectiveness, and perceived opportunities, respectively.

Strategy use free-response questions

At the 2 and 6 week explicit measure assessments, the intervention group also completed questionnaires in which they gave open-ended responses about their experiences using the strategies. For each strategy, the participants were asked whether they had used the strategy since their last in-lab session. If they had used a strategy, the participants were subsequently asked to describe one or two situations in which they had used the strategy and to provide general comments about their experiences using the strategy. At the end of the questionnaire, the participants were asked to share any additional comments about implementing the strategies.

Procedure

During the first lab session, all participants completed the IAT and received feedback about their performance. Specifically, an experimenter calculated the participants' IAT scores and asked them to type their scores into the computer. The computer provided participants with an interpretation of their IAT performance based on their *D*-scores, saying that they had a strong preference for Blacks over Whites (*D*-score less than or equal to −.65), a moderate preference for Blacks over Whites (*D*-score between −.65 and −.35), a slight preference for Blacks over Whites (*D*-score between −.35 and −.15) no preference for Whites or Blacks (*D*-score between −.15 and .15), a slight preference for Whites over Blacks (*D*-score between .15 and .35), a moderate preference for Whites over Blacks (*D*-score between .35 and .65) or a strong preference for Whites over Blacks (*D*-score over .65).

To the extent that the participants implicitly favored White people over Black people (as 90% of our participants did), we expected that being confronted with evidence of this bias would increase participants' awareness of their bias (Monteith, Voils, & Ashburn-Nardo, 2001). After receiving feedback, control group participants were dismissed; they were reminded, however, that they would return to the lab at two subsequent points in time and would receive questionnaires to fill out between their lab sessions. People in the intervention group were presented with a 45-minute narrated and interactive slideshow separated into education and training sections.[3] The education section introduced the idea of prejudice as a habit, as well as how implicit biases develop and are automatically activated without intention. This section also explained the general logic behind how the IAT measures implicit bias (without giving the participants a specific strategy to "beat the IAT"; Kim, 2003) and dispelled alternate explanations for IAT bias (e.g., the order of the congruent and incongruent trials, color associations). Participants were then taught about the prevalence of implicit race biases and how they can lead people to unwittingly perpetuate discrimination. Specifically, participants learned about research linking IAT bias to a wide range of discriminatory outcomes in domains such as health, employment, and everyday interpersonal interactions.

Strategies for reducing implicit race bias

The training section provided participants with a list of five strategies culled from the literature and adapted for the intervention (see Fig. 1). The program explained the strategies in straightforward language with concrete examples of everyday situations in which they could be used. Participants were then asked to generate situations in which they could use each strategy. Participants were told that although none of the strategies are difficult to implement, each requires some effort. In addition, the program emphasized how the strategies (explained below) are mutually reinforcing. For example, contact with counter-stereotypic others provides grist for counter-stereotypic imaging as well as providing opportunities for individuation, perspective taking and stereotype replacement. Similarly, perspective taking can enhance stereotype replacement and individuation by encouraging people to see the world from the eyes of a stigmatized other. As a set, the strategies were offered as a powerful toolkit for breaking the prejudice habit. The program also stressed that practicing the strategies would help them to reduce implicit bias and, hence, break the prejudice habit. Following the education and training sessions, participants were reminded that they would return to the lab for two subsequent sessions and would receive questionnaires to complete between the lab sessions. Participants were then dismissed.

Stereotype replacement

This strategy involves replacing stereotypical responses for non-stereotypical responses. Using this strategy to address personal stereotyping involves recognizing that a response is based on stereotypes, labeling the response as stereotypical, and reflecting on why the response occurred. Next one considers how the biased response could be avoided in the future and replaces it with an unbiased response (Monteith, 1993). A parallel process can be applied to societal (e.g., media) stereotyping.

Counter-stereotypic imaging

This strategy involves imagining in detail counter-stereotypic others (Blair et al., 2001). These others can be abstract (e.g., smart Black

[3] Materials available upon request to the first author.

people), famous (e.g., Barack Obama), or non-famous (e.g., a personal friend). The strategy makes positive exemplars salient and accessible when challenging a stereotype's validity.

Individuation

This strategy relies on preventing stereotypic inferences by obtaining specific information about group members (Brewer, 1988; Fiske & Neuberg, 1990). Using this strategy helps people evaluate members of the target group based on personal, rather than group-based, attributes.

Perspective taking

This strategy involves taking the perspective in the first person of a member of a stereotyped group. Perspective taking increases psychological closeness to the stigmatized group, which ameliorates automatic group-based evaluations (Galinsky & Moskowitz, 2000).

Increasing opportunities for contact

This strategy involves seeking opportunities to encounter and engage in positive interactions with out-group members. Increased contact can ameliorate implicit bias through a wide variety of mechanisms, including altering the cognitive representations of the group or by directly improving evaluations of the group (Pettigrew, 1998; Pettigrew & Tropp, 2006).

Results

Data analytic plan

The intervention and control groups did not differ on any of the measures at baseline, all $ps \geq .50$ (see Table 1; for correlations between the study variables within each condition, see Table 2). All of the analyses were conducted using a series of General Linear Models (GLMs) using the baseline measurement of the dependent variable as a covariate, an approach that is more powerful than using a difference score approach when there are no pre-test differences on the dependent variable (Van Breukelen, 2006). Predictors were centered prior to testing interactions. All missing data were handled through multiple imputation (Rubin, 1987).[4]

To examine the multifarious effects of the intervention after baseline, we conducted a series of analyses. First, we analyzed the effect of the intervention on our implicit measure of race bias. We then analyzed the effect of the intervention on each of our explicit measures. Next, we conducted a series of moderation analyses of the impact of the intervention on implicit bias using the baseline and week 2 measurements of each of the explicit measures. Finally, we conducted a series of analyses within the intervention group to determine whether either the post-manipulation reactions to the strategies or the coded variables from the free response reports of strategy use predicted reductions in implicit bias.

Effect of the intervention on implicit race bias

A major goal of this study was to examine the impact of the intervention on the magnitude of implicit race bias assessed over time. As shown in Fig. 2, the intervention was successful. Following the manipulation, intervention group participants had lower IAT scores than control group participants, $B = -.19$, $t(88) = -2.82$, $p = .006$, $\Delta R^2 = .081$. Moreover, the effects of the intervention on implicit race bias at 4 and 8 weeks were not systematically different from each other, $B = .091$, $t(88) = .82$, $p = .42$, $\Delta R^2 = .008$, indicating that the reduction in implicit race bias persisted throughout the 8-week interval. These data provide the first evidence that a controlled, randomized intervention can produce enduring reductions in implicit bias.[5]

Table 1
Means and standard deviations of implicit and explicit variable by condition.

		Intervention condition ($N = 53$)		Control condition ($N = 38$)	
		Mean	*SD*	*Mean*	*SD*
Baseline	IAT *D*-score	0.47	0.42	0.42	0.36
	ATB	5.67	0.66	5.57	0.72
	EMS	5.37	1.80	5.52	1.75
	IMS	7.64	0.97	7.47	1.41
	Shoulds	2.03	0.65	1.91	0.72
	Woulds	3.16	1.00	3.20	0.88
	Discrepancies	1.17	0.90	1.29	0.80
	Concern	6.08	1.06	6.05	1.37
Time 2	IAT *D*-score	0.32	0.41	0.54	0.36
	ATB	5.52	0.71	5.58	0.72
	EMS	5.02	1.87	5.52	1.75
	IMS	7.11	1.19	7.47	1.41
	Shoulds	1.90	0.55	1.99	0.78
	Woulds	3.51	1.04	3.29	0.94
	Discrepancies	1.63	1.07	1.34	1.06
	Concern	6.42	0.95	6.24	1.20
Time 3	IAT *D*-score	0.30	0.42	0.47	0.41
	ATB	5.60	0.73	5.53	0.72
	EMS	5.00	1.64	5.21	1.95
	IMS	7.28	1.28	7.28	1.43
	Shoulds	1.89	0.54	1.90	0.67
	Woulds	3.38	0.93	3.15	0.85
	Discrepancies	1.49	0.86	1.26	0.76
	Concern	6.57	1.04	5.86	1.24

Note: None of the measures differ by condition at baseline.

Effect of the intervention on the explicit measures

We were also interested in whether the intervention affected explicit variables previously argued to be important to the prejudice reduction process. The intervention manipulation created no changes in either the participants' reported racial attitudes or their internal/external motivations to respond without prejudice (all $ps \geq .53$). It did, however, affect participants' concern about discrimination and their awareness of their personal bias as revealed through their prejudice-relevant discrepancies. Intervention group participants had higher concern than control group participants following the intervention manipulation, $B = .43$, $t(88) = 2.24$, $p = .028$, $\Delta R^2 = .049$. As shown in Fig. 3, participants in the control group had a constant level of concern throughout the duration of the study, $B = -.044$, $t(89) = -.22$, $p = .83$, $d = -.030$, whereas participants in the intervention group increased in concern after they received the intervention, $B = .40$, $t(89) = 2.10$, $p = .042$, $d = .38$. The effect of the intervention manipulation on concern also grew more pronounced over time, $B = .54$, $t(88) = 2.35$, $p = .021$, $\Delta R^2 = .061$. We explored whether the increases in concern observed in the intervention group predicted the decreases in implicit race bias, but this was not the case, $B = -.033$, $t(88) = -.60$, $p = .55$, $\Delta R^2 = .005$, suggesting that the decreases in implicit bias and increases in concern were somewhat independent effects of the intervention.

[4] We used five imputations, each of which were iterated until convergence. All results were pooled Rubin's rules. Although we report unadjusted degrees of freedom, all other reported statistics are adjusted for the amount of information lost due to missingness.

[5] An anonymous reviewer brought up the possibility that the participants in the training condition spontaneously discovered strategies to "beat the IAT" after they received information on how the IAT measures implicit bias (Kim, 2003). To guard against this possibility, the reviewer suggested analyzing response latencies on the congruent and incongruent trials separately to determine whether the participants were deliberately slowing their responses on the congruent trials. After the intervention was administered, participants in the training and control groups had equal latencies on both congruent trials, $B = 18.75$, $t(88) = .76$, $p = .45$, $\Delta R^2 = .006$, and on incongruent trials, $B = -34.24$, $t(88) = -1.02$, $p = .31$, $\Delta R^2 = .011$.

Table 2
Correlations between the major study variables within the control and training conditions. Correlations in the control condition are shown below the diagonal, whereas correlations within the training condition are shown above the diagonal.

	IAT 1	IAT 2	IAT 3	ATB 1	ATB 2	ATB 3	EMS 1	EMS 2	EMS 3	IMS 1	IMS 2	IMS 3	Should 1	Should 2	Should 3	Would 1	Would 2	Would 3	Discrepancies 1	Discrepancies 2	Discrepancies 3	Concern 1	Concern 2	Concern 3
IAT 1	–	0.36	0.01	−0.14	−0.20	−0.15	0.12	0.16	0.13	−0.05	0.08	0.16	0.09	0.08	0.05	0.15	0.25	0.24	0.11	0.20	0.22	−0.07	−0.08	0.11
IAT 2	0.33	–	0.27	−0.27	−0.38	−0.35	0.21	0.31	0.25	−0.22	−0.23	−0.19	0.26	0.24	0.24	0.33	0.31	0.31	0.19	0.19	0.18	−0.28	−0.36	−0.10
IAT 3	0.21	0.18	–	0.01	−0.09	−0.05	0.03	0.14	0.09	0.09	−0.01	0.11	−0.13	−0.11	−0.10	−0.03	0.00	−0.10	0.06	0.05	−0.04	−0.03	−0.34	−0.1
1ATB 1	−0.15	−0.17	0.02	–	0.74	0.75	−0.16	−0.03	−0.14	0.37	0.48	0.44	−0.65	−0.58	−0.66	−0.58	−0.43	−0.44	−0.19	−0.13	−0.04	0.42	0.30	0.42
ATB 2	−0.20	−0.19	0.01	0.77	–	0.83	−0.22	−0.05	−0.12	0.34	0.63	0.50	−0.46	−0.57	−0.66	−0.54	−0.59	−0.55	−0.29	−0.30	−0.16	0.31	0.42	0.48
ATB 3	−0.08	−0.19	0.00	0.73	0.68	–	−0.24	−0.15	−0.27	0.35	0.60	0.48	−0.51	−0.64	−0.71	−0.62	−0.57	−0.64	−0.35	−0.24	−0.22	0.40	0.46	0.53
EMS 1	0.18	−0.01	−0.16	−0.04	−0.03	−0.10	–	0.43	0.56	−0.01	−0.11	−0.25	0.15	0.15	0.05	0.37	0.19	0.14	0.33	0.11	0.11	−0.18	−0.16	−0.15
EMS 2	0.01	−0.04	0.02	0.06	0.07	0.05	0.46	–	0.66	−0.07	0.24	0.01	0.09	−0.05	−0.06	0.19	0.39	0.31	0.16	0.40	0.38	−0.23	−0.21	−0.03
EMS 3	−0.08	−0.09	0.02	−0.05	−0.08	−0.11	0.37	0.67	–	−0.14	0.14	0.08	0.15	0.02	−0.06	0.36	0.37	0.25	0.31	0.35	0.31	−0.31	−0.40	−0.17
IMS 1	−0.04	−0.09	0.16	0.64	0.69	0.32	0.04	0.10	−0.10	–	0.31	0.44-	0.41	−0.40	−0.23	−0.39	−0.36	−41	−0.15	−0.16	−0.30	0.28	0.11	0.12
IMS 2	−0.04	−0.20	0.12	0.39	0.58	0.30	0.11	0.30	−0.08	0.75	–	0.77	−0.44	−0.63	−0.64	−0.32	−0.24	−0.32	−0.04	0.07	0.08	0.16	0.33	0.42
IMS 3	−0.19	−0.27	−0.04	0.58	0.81	0.48	0.00	0.24	0.00	0.77	0.80	–	−0.49	−0.60	−0.63	−0.28	−0.22	−0.28	0.05	0.08	0.11	0.25	0.16	0.37
Should 1	0.23	0.26	−0.05	−0.51	−0.49	−0.55	−0.11	−0.25	−0.08	−0.40	−0.46	−0.51	–	0.66	0.58	0.57	0.35	0.40	−0.10	0.02	0.05	−0.30	−0.19	−0.23
Should 2	0.18	0.37	0.08	−0.48	−0.56	−0.55	−0.22	−0.35	−0.21	−0.44	−0.58	−0.64	0.70	–	0.75	0.47	0.20	0.32	0.05	−0.29	−0.15	−0.20	−0.26	−0.36
Should 3	0.38	0.38	0.16	−0.58	−0.71	−0.61	−0.14	−0.38	−0.20	−0.38	−0.49	−0.65	0.66	0.73	–	0.32	0.37	0.44	−0.07	−0.01	−0.19	−0.31	−0.26	−0.48
Would 1	0.05	0.15	0.08	−0.61	−0.50	−0.61	0.07	−0.01	0.19	−0.26	−0.35	−0.39	0.38	0.33	0.48	–	0.62	0.65	0.76	0.37	0.50	−0.22	−0.08	−0.16
Would 2	0.13	0.19	−0.12	−0.49	−0.64	−0.40	0.11	0.28	0.21	−0.48	−0.35	−0.49	0.19	0.21	0.32	0.51	–	0.85	0.47	0.88	0.69	−0.25	−0.25	−0.20
Would 3	0.05	0.31	0.01	−0.69	−0.74	−0.64	0.17	0.11	0.16	−0.57	−0.40	−0.66	0.34	0.41	0.56	0.62	0.76	–	0.48	0.67	0.80	−0.18	−0.03	−0.12
Discrepancies 1	−0.14	−0.06	0.11	−0.19	−0.09	−0.16	0.16	0.19	0.25	0.07	0.03	0.03	−0.43	−0.24	−0.06	0.67	0.35	0.34	–	0.44	0.57	−0.03	0.05	−0.01
Discrepancies 2	−0.01	−0.10	−0.16	−0.08	−0.15	0.04	0.25	0.49	0.33	−0.10	0.11	0.03	−0.33	−0.53	−0.24	0.21	0.72	0.37	0.47	–	0.74	−0.15	−0.12	−0.02
Discrepancies 3	−0.29	0.01	−0.13	−0.26	−0.21	−0.18	0.33	0.48	0.37	−0.31	−0.02	−0.16	−0.21	−0.19	−0.27	0.28	0.59	0.65	0.44	0.65	–	0.01	0.14	0.19
Concern 1	0.15	0.18	0.05	0.53	0.48	0.26	0.07	0.37	0.23	0.54	0.46	0.46	−0.12	−.21	−0.31	−0.45	−0.25	−0.34	−0.34	−0.07	−0.10	–	0.39	0.27
Concern 2	0.00	0.10	0.07	0.23	0.56	0.32	0.05	0.25	0.03	0.52	0.55	0.60	−0.31	−0.26	−0.42	−0.17	−0.27	−0.36	0.07	−0.05	−0.03	0.41	–	0.53
Concern 3	−0.07	0.14	0.06	0.33	0.53	0.49	−0.04	0.21	0.09	0.36	0.31	0.57	−0.27	−0.34	−0.41	−0.17	−0.23	−0.41	0.05	0.04	−0.10	0.29	0.66	–

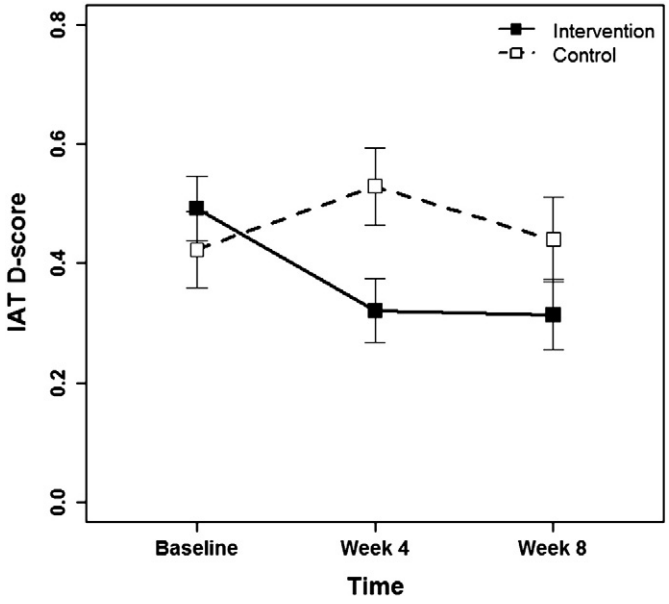

Fig. 2. IAT *D*-scores for intervention and control group participants before the manipulation and 4 and 8 weeks after the manipulation. Higher numbers indicate higher levels of implicit bias. IAT *D*-scores did not differ before the manipulation, but after the manipulation, participants who received the intervention had lower IAT scores than participants who did not. Error bars represent ±1 standard error of the GLM point estimate.

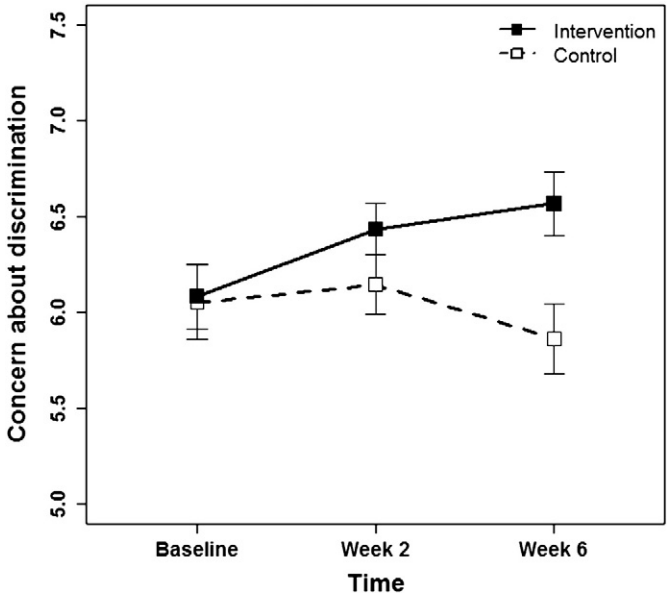

Fig. 3. Concern about discrimination by condition 4 weeks before the manipulation and 2 and 6 weeks after the manipulation. Higher numbers indicate higher levels of concern. Concern did not differ by condition before the manipulation, but after the manipulation, participants who received the intervention were more concerned about discrimination than participants who did not. Error bars represent ±1 standard error of the GLM point estimate.

Compared to control group participants, intervention group participants reported greater discrepancies between their *should*s and *would*s following the intervention, $B = .38$, $t(88) = 2.31$, $p = .024$, $\Delta R^2 = .047$. Separate analyses on the *should* and *would* indices revealed that although the standards for responding towards Black people (*should*s) remained unchanged in the intervention group relative to the control group, $B = -.095$, $t(88) = -.93$, $p = .36$, $\Delta R^2 = .005$, participants in the intervention group increased in how much they predicted they would respond with bias in intergroup situations (*would*s) following the intervention manipulation, $B = .29$, $t(88) = 1.93$, $p = .057$, $\Delta R^2 = .026$. As a set, the analyses on *should*s, *would*s, and prejudice-relevant discrepancies suggest that the intervention caused people to become more aware of their personal bias, while leaving people's standards for behavior in prejudice-relevant situations unchanged. As with concern, we explored whether the increases in discrepancies predicted the decreases in implicit race bias, but this was not the case, $B = -.037$, $t(88) = -.67$, $p = .50$, $\Delta R^2 = .005$. The effect of the intervention on discrepancies was not systematically different at week 4 and week 8, $B = -.067$, $t(88) = -.37$, $p = .72$, $\Delta R^2 = .004$ (see Fig. 4).

Moderation analyses

We next investigated whether any of the explicit measurements taken at baseline or week 2 moderated the effect of the intervention on implicit bias. Somewhat surprisingly, none of the explicit measures taken at baseline moderated the effect of the intervention on implicit bias, all $ps \geq .15$.[6] The only week 2 explicit variable to emerge as a moderator was concern about discrimination, $B = -.15$, $t(86) = -2.46$, $p = .016$, $\Delta R^2 = .058$, all other $ps \geq .23$. As shown in Fig. 5, intervention condition participants with more concern about discrimination at week 2 had particularly low levels of implicit bias at weeks 4 and 8. This effect remained from week 4 to week 8, $B = -.043$, $t(86) = -.082$, $p = .94$, $\Delta R^2 = .002$, indicating that people high in concern about discrimination at week 2 retained the reductions in IAT bias 8 weeks after the intervention. The interactive effect of condition and concern was entirely driven by a robust relationship between concern and implicit bias in the intervention condition, $B = -.16$, $t(86) = -2.90$, $p = .009$, $\Delta R^2 = .12$, a relationship that was entirely absent in the control condition, $B = .028$, $t(86) = .55$, $p = .59$, $\Delta R^2 = .005$.

We believe that these interactive effects of the intervention manipulation and concern on implicit race bias are of large practical significance. Compared to high concern control group participants, the predicted IAT scores of high concern participants in the intervention

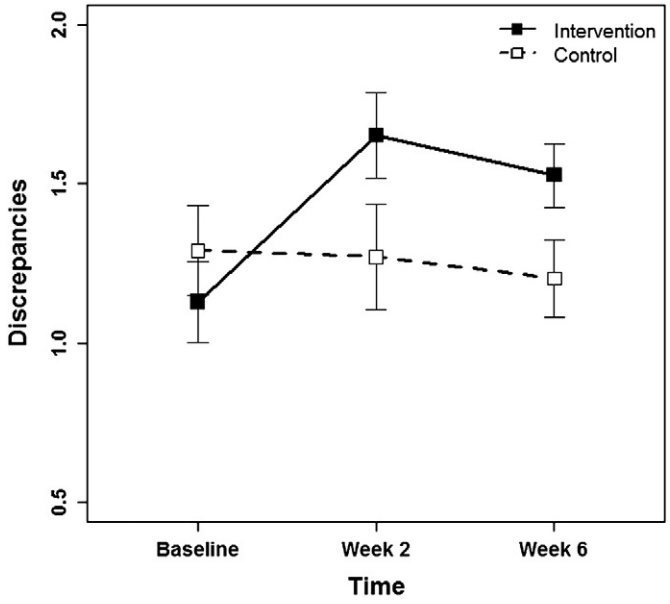

Fig. 4. Discrepancies between self-reported standards (*should*s) and predicted actual reactions (*would*s) to Blacks by condition 4 weeks before the manipulation and 2 and 6 weeks after the manipulation. Higher numbers indicate that participants predict they would react to Blacks with more bias than they believe is appropriate. Discrepancies did not differ by condition before the manipulation, but after the manipulation, participants who received the intervention had larger discrepancies than participants who did not. Error bars represent ±1 standard error of the GLM point estimate.

[6] The three-way interaction between internal motivation, external motivation, and condition was also not significant, $B = -.028$, $t(82) = -.67$, $p = .52$, $\Delta R^2 = .023$.

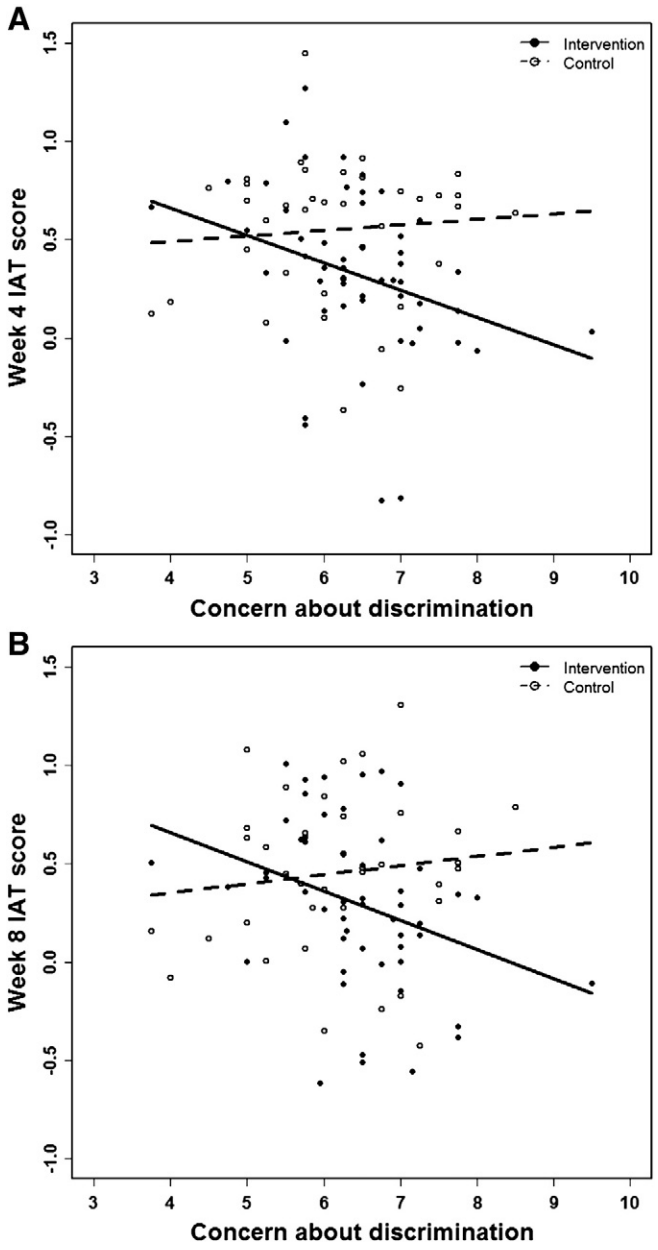

Fig. 5. Week 2 concern about discrimination plotted against week 4 (Panel A) and week 8 (Panel B) IAT *D*-scores with prediction lines from the GLM. Higher numbers indicate higher levels of implicit bias and greater levels of concern. Prediction lines are plotted at ±1standard deviation from the mean on concern. Within the intervention condition, concern was associated with lower IAT scores at weeks 4 and 8. Within the control condition, concern was unrelated to IAT scores.

group were .38 units lower at week 4 and .31 units lower at week 8. These decreases are, for example, large enough to bring someone from the "moderate preference for Whites over Blacks" feedback category (*D*-score between .35 and .65) down into the "slight preference for Whites over Blacks" feedback category (*D*-score between .15 and .35). The interaction remained significant when controlling for all the other explicit measures, $B=-.14$, $t(82)=-2.31$, $p=.024$, $\Delta R^2=.049$.

Perceptions of strategies analyses

To gain some understanding about what the participants in the intervention condition did to reduce their implicit race bias, we examined whether, after controlling for baseline implicit bias, the participants' reactions to the strategies were associated with lower IAT bias at week 4 and week 8. These analyses revealed that self-reported likelihood to use the strategies was associated with lower implicit race bias, $B=-.14$, $t(50)=-3.18$, $p=.003$, $\Delta R^2=.17$, an effect that was not systematically different between week 4 and week 8, $B=-.050$, $t(50)=-.69$, $p=.50$, $\Delta R^2=.014$. Though the effects were weaker, perceived opportunities also emerged as a predictor of lower implicit race bias, $B=-.092$, $t(50)=-2.40$, $p=.021$, $\Delta R^2=.10$, as did perceived effectiveness, $B=-.12$, $t(50)=-1.86$, $p=.069$, $\Delta R^2=.063$. Perceived opportunities was, however, highly correlated with likelihood, $r=.76$, as was effectiveness, $r=.40$, and when all three variables were allowed to predict post-baseline implicit race bias, only the relationship with likelihood remained robust, $B=-.13$, $t(48)=-2.03$, $p=.051$, $\Delta R^2=.078$. Thus, to have a reduction in implicit race bias, it appears that participants had not only to perceive opportunities to implement the strategies and view the strategies as effective, but also to believe that they were likely to use them. Perceived difficulty of implementing the strategies and willingness to use the strategies did not emerge as predictors of reduced bias, $ps>.23$.

Strategy use free-response analyses

Participants' descriptions of their strategy use were rich and complex. To capture this complexity while maintaining objectivity, we used a text mining approach to calculate frequencies of theoretically important word stems in the descriptions. The overarching goal of this approach was to determine if the constructs thought to be important in the habit-breaking model were included in participants' descriptions, and, if so, if they were also related to reductions in implicit bias.

To that end, we used the tm package for *R* (Feinerer, Hornik, & Meyer, 2008) to load the participants' responses from each time point into a computerized corpus of responses. Each response was screened for the default stop words from the tm package (e.g., "the", "a", "and"), and the resulting sets of words were reduced to word stems. We then chose a standard psycholinguistic dictionary (WordNet; Fellbaum, 1998) to look up synonyms of words that related to the three theoretically important categories in the prejudice habit model: motivation (or the decision that prejudice is wrong), awareness, and the implementation of strategies to combat bias. After validating that the meanings of all the resulting words matched that of the target categories, we eliminated all words that had not been used by at least two people at each time point. The resulting word stems for each category, as well as the means and standard deviations of their frequency of use, are displayed in Table 3. Sample participant responses with bolded target words are displayed in Table 4.

Table 3
Means and standard deviations of word stem frequencies from the free-response answers concerning strategy use, broken into three different conceptual categories.

		Week 2		Week 6	
		Mean	*SD*	*Mean*	*SD*
Awareness stems	awar	0.15	0.50	0.09	0.35
	realiz	0.92	1.02	0.81	0.88
	recogn	0.25	0.59	0.11	0.32
	understand	0.17	0.58	0.09	0.30
	Total	1.49	1.30	1.11	1.12
Motivation stems	wrong	0.26	0.56	0.15	0.36
	unfair	0.38	0.66	0.09	0.40
	Total	0.64	0.98	0.25	0.65
Implementation stems	implement	0.49	0.89	0.87	1.49
	practic	0.11	0.32	0.21	0.53
	appli	0.26	0.65	0.30	0.72
	use	0.45	0.93	0.91	1.29
	tri	0.64	0.94	0.40	0.77
	Total	1.96	2.09	2.68	2.67

Table 4
Sample participant responses to the free-response questions regarding strategy use and their corresponding word stem frequencies for each conceptual category. Word stems included in the analyses are highlighted in bold.

Category	Frequency	Response
Awareness	3	I was at a house party one weekend when two tall and strong African Americans walked in. I immediately assumed they must be on the football or some other sports team. I realized that thought was stereotypical and decided that such an assumption should not be made. It is very useful in evaluating one's thoughts. Two of my favorite shows are CSI and Law and Order: SVU. In many instances I was able to implement this technique. I could pick out stereotypes and realize they were unfair. I think this technique is very useful because it shows that discrimination is not something a person may just do personally, but the media discriminates as well. Two of my cousins are
Motivation	2	African American, so it is easy to implement this technique, especially when racist jokes are made. It is very useful for people who have someone in their life that does not prove the stereotypes true. At the party I described above I chose to talk to the guys and get to know about them. It ended up that both were indeed on the football team but
Implementation	8	I found that there was so much more to them then that, that single fact no longer seemed important. Although not always possible, this technique is useful when it can be used. I implemented this technique during the same party situation, I realized how unfair I would have been to be seen and appreciated for only my athletic abilities. I think this technique is best used following a situation, as a post analysis. I don't always remember or think about using the techniques.
Awareness	2	Before spring break I was riding the bus and an older black gentlemen sat in the seat next to me. I was about to sit closer to the wall but then realized that this was a stereotypical response and stayed where I was. This technique is very effective and very easy to implement in daily activities. I had to go get a drug test for my new job. While I was in the clinic a black woman walked up to the desk and the receptionist assumed she was there for a drug test also. I thought that was stereotyping on her part and I was right. The woman was there because she was injured at work. This technique is easy to recognize in a society filled with stereotypes and prejudice. Over break I
Motivation	0	was driving to work and a black woman flew by me on 151. I immediately figured she was rushing to work and was late all the time. Really I had no clue what the situation was. This could be her first time being late to work, or maybe she just likes to speed. I didn't know the individual so I was simply generalizing. This process requires a little bit of thought but it is good to look at people as individuals and not lump people together. This process seems like it would be very difficult because I have never been black and have no clue how hard it would be to grow up dealing with stereotypes and prejudice. I recently got a job at an employer that actually believes in equal
Implementation	5	opportunity and I have many black co workers. I have decided to try to get to know my black coworkers better since at my last job they did not really hire black people and I had no opportunity for this. This technique requires the most physical work but should be easy if stereotypes are put aside. 1, 2, 4 are quite natural to do. 3, 5, and 6 are sometimes unnatural and take a little more work to implement. Although it's not always easy using these techniques they are a good way to work at getting rid of stereotypes we may have learned from personal experience or the media. It would have been nice if we would have been told to start implementing these techniques immediately following session 1.
Awareness	3	The movie showed a black male with an afro, blowing things up and stealing things. I realize this is stereotypical and replace it. I don't feel like this type of action in the media creates much impact on a racial level. I was walking down an alley when a black man approached me and my first response was to stay as far away from him as possible because I was afraid he would mug me seeing that I had cash in my hand. I realized this man was homeless and not scary at all but rather trying to get some food. This is one of the easy techniques to apply because it is quite easy to establish that something is stereotypical and alter the way you think about the situation. I recently had a party and a black male showed up, I was hesitant to let him wander around in my house because I didn't know him that well but I had no issue with the same situation
Motivation	0	white males. Instead I talked with him for nearly 30 minutes and realized I had nothing to worry about. This technique comes across fairly easy if the situation presents itself to you. I recently went to a dance event sponsored by a group majority of black students. This is an easy technique to gain knowledge of a different race or religious
Implementation	2	view. Increasing opportunities Individuating. It doesn't seem that difficult but I don't feel like it changes my subconscious attitudes towards blacks.
Awareness	1	Scanning IDs at the Nat, on spring break in myrtle beach I realized I was thinking stereotypical thoughts when I saw a black person walking down the street and was
Motivation	0	afraid, so I remembered all the times this happened and nothing happened. Upon thinking a black person might not be as smart as others in the class I imagined what that
Implementation	0	would feel like if people thought of me like that. I think it is a good way to prevent stereotypical thoughts.

We then fit a series of models to determine whether, after controlling for baseline implicit bias, word stem use from each conceptual category at week 2 was related to implicit bias at week 4 or week 8 and whether word stem use at week 6 related to implicit bias at week 8. The only significant effect to emerge from these analyses was that more frequent use of implementation-related word stems at week 2 predicted reduced implicit bias at week 4, $B = -.068$, $t(50) = -2.47$, $p = .017$, $\Delta R^2 = .11$, all other $ps > .21$. We also tested whether likelihood was related to use of implementation-related word stems, but, in fact, these two variables were almost entirely independent of each other, $r = .027$. When used to simultaneously predict week 4 implicit bias, both likelihood and implementation-related word use were related to decreased implicit bias, $B = -.13$, $t(49) = -2.40$, $p = .028$, $\Delta R^2 = .12$, and $B = -.068$, $t(49) = -2.66$, $p = .012$, $\Delta R^2 = .11$, respectively. Likelihood and week 2 implementation-related word stem use jointly accounted for fully 23% of the variance in week 4 implicit bias.

Discussion

Overall, our results provide compelling and encouraging evidence for the effectiveness of our multifaceted intervention in promoting enduring reductions in implicit bias. As such, this study provides a resounding response to the clarion call for methods to reduce implicit bias and thereby reduce the pernicious, unintended discrimination that arises from implicit biases. Reductions in implicit bias that emerged by week 4 following the intervention persisted to week 8. Such enduring reductions in implicit bias following a bias reduction intervention are unprecedented in the literature (Paluck & Green, 2009). Although some previous research has established that people who choose to immerse themselves in a context either rich in counter-stereotypic exemplars (e.g., a women's college; Dasgupta & Asgari, 2004) or that is conducive to the regular discussion of issues related to implicit bias (e.g., in a course on stereotyping and prejudice; Rudman, Ashmore, & Gary, 2001) show reduced implicit bias, our study is the first to our knowledge to produce long-term change in implicit bias using a randomized, controlled design.

Another encouraging finding was the simultaneous effect of the intervention on increasing people's self-reported concern about discrimination and prejudice-relevant discrepancies. The intervention thus seems to increase both personal awareness of one's bias and a general concern about discrimination in society. The effect of the intervention on concern also grew more pronounced over time, potentially suggesting that the intervention created an increased caring about subtle instances of bias and discrimination. We suspect that the intervention caused people to become more attuned to their own spontaneous biases and everyday instances of discrimination and that these experiences, coupled with increased caring, may have created ever-rising levels of concern. Future studies should increase the frequency of measurement of both concern and discrepancies to determine the precise time-course of the changes on these variables. Such studies should also measure concern and discrepancies immediately after the administration of the intervention to determine whether the impact on these variables occurs immediately after the intervention or only after people have had time to observe subtle discrimination within themselves and in their environment.

Interestingly, none of the explicit variables measured at baseline served as moderators of the effect of the intervention on implicit bias. This is somewhat surprising given that some of the variables tested, such as internal and external motivations to respond without prejudice and prejudice-relevant discrepancies, have been previously implicated in various processes that should affect receptivity to bias-reducing interventions (Monteith, 1993; Plant & Devine, 2009). We speculate that our intervention, which was both interactive and narrated, created little opportunity for unmotivated or unaware participants to tune out the education and training components. Hence, people who may not otherwise have been engaged by the intervention may have, despite the lack of a priori personal or external motivation to respond without prejudice, found themselves compelled by the content of the intervention and therefore began to make efforts to regulate their bias.

In contrast to the baseline explicit measures, one explicit measure at collected at week 2, concern about discrimination, did emerge as a moderator of the intervention's effect. This finding crucially implicates concern about discrimination in the bias-reducing process. Given that week 2 concern, and not baseline concern, moderated the effect of the intervention, this finding also suggests that concern is not important when bias-reducing strategies are learned, but that it is important afterwards, when people become aware of personal or societal expressions of bias and must translate their knowledge of bias-reducing strategies into action. Given the importance of concern in predicting who reduced their bias, this finding highlights the need to explore what aspects of the multifaceted intervention were responsible for increasing concern. Education might be essential for evoking concern, but other components of the intervention may be necessary as well (e.g., being situationally aware of one's implicit bias prior to the narrated slide show). Exploring these issues will enable the design of more effective interventions in the future and help us understand the precise psychological process that implicates concern in bias reduction.

Our findings regarding the free-response descriptions of strategy use and reported likelihood of use suggest a potential process responsible for the initial reduction in implicit bias and the maintenance of that reduction — use and anticipated use of the strategies. The use of implementation-related word stems in describing strategy use at week 2, but not awareness or motivationally related word stems, predicted reduced implicit bias at week 4. This suggests that the use and practice of strategies in the period immediately following the intervention is particularly important to initial bias reduction. The fact that word stem usage at week 2 did not predict implicit bias at week 8 suggests that the factors involved in initial strategy deployment are different from the factors involved in the maintenance of decreased implicit bias. Likelihood predicted reduction in implicit bias at both week 4 and week 8, even after controlling for all the other strategy measures. This suggests that the intervention generates intentions to use the strategies that are crucial to the maintenance of the bias-reducing process. Future work should continue to explore how and under what circumstances the strategies are integrated into people's lives to effect change in implicit bias.

The word frequency findings, which implicated overall strategy use as being important for reducing implicit bias, do not reveal whether people generally prefer one strategy over another or whether one particular strategy is more effective than another at producing the various outcomes of the intervention. Because the different strategies exert their effects through different psychological mechanisms, and because the strategies are likely used in different situations, they might have specialized effects on outcomes relevant to the regulation of implicit bias. For example, the stereotype replacement technique requires becoming situationally aware of the fact that one has or is likely to have a biased response. Consequently, frequent use of the stereotype replacement technique might lead to increased discrepancies as people become chronically sensitive into the fact that they respond with stereotypic biases. In contrast, the perspective taking technique requires experiencing the world from the perspective of a stigmatized person. As people use this technique, they might come to better understand the consequences of subtle discrimination for outgroups, thereby becoming more concerned about discrimination. Unfortunately, the present study did not include precise quantitative indicators of use of the five strategies taught to the participants. Collecting more precise indicators of strategy use may help to delineate the extent to which use of specific strategies relates to specific outcomes and may shed light on the mechanisms underlying the regulation of implicit bias.

The word frequency and likelihood findings, combined with the findings about change in implicit bias, personal awareness, and concern, support the prejudice habit model and other dual-process models that

identify effort as necessary for implicit bias reduction (Devine, 1989; Devine & Monteith, 1993; Smith & DeCoster, 2000; Strack & Deutsch, 2004). The support for the prejudice habit model could be strengthened through obtaining measures of the specific behavioral process (e.g., use of a particular strategy or set of strategies) required to produce change in implicit bias, concern, and personal awareness. Nevertheless, we believe that our findings provide an important demonstration that an intervention can engage the long-term regulation of implicit bias, as well as some preliminary evidence that the regulation occurs through the use of strategies.

Future studies will need to establish the specific behavioral, cognitive, affective, and neural mechanisms through which this intervention exerts its effects. In addition to measuring when, where, and with what frequencies people use the various strategies, future studies could, for example, use multinomial modeling (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Payne, 2001) and behavioral neuroscience (Cunningham et al., 2004; Phelps et al., 2000) to determine whether the intervention changed negative associations about Black people or led to more efficient control over automatic bias. The identification of specific mechanisms will lead to a better understanding of how, why, and under what circumstances the intervention will be effective.

Our intervention was multifaceted by design. This decision was guided by the fact that (1) we wished to test whether it was possible to engage a complex self-regulatory process involved in voluntary efforts to decrease implicit bias over time and (2) it was not possible to specify a priori which elements of the intervention (e.g., feedback regarding one's personal level of implicit bias, education about the nature and consequences of implicit bias, training regarding strategies to reduce bias and opportunities to report on strategy use, questionnaires that subtly remind the participants about the material presented in the intervention) would be necessary or sufficient to engage the regulatory process (Howell, 1982; Kaufman, 2003; Prochaska & Velicer, 1997). The various components of the intervention were intended to increase awareness of bias, increase concern about discrimination, and teach strategies that reduce bias as well as assess strategy use.

Though effective overall, the complexity of the intervention results in ambiguity regarding which components are responsible for its various effects. For example, as shown in our control group, feedback about the presence of implicit bias is not sufficient to trigger bias reduction, but it may necessary. Education may play a specialized role in increasing awareness and concern, but both education and training may be necessary to produce changes in implicit bias. Additionally, our effort to assess strategy use between the laboratory sessions may have been crucial for the intervention's effectiveness by stimulating participants to think about the strategies and how they could be applied in their everyday experiences.

Although the complexity of the intervention brings ambiguity in the interpretation of the effects of the intervention, it is also likely that there is no single “magic bullet” that, by itself, prompts the regulation of implicit bias and the multifarious changes in concern and awareness such self-regulation brings. Instead, several components likely work in combination to prompt situational awareness of one's bias and translate that awareness into chronic awareness, concern, and self-regulatory effort. Future studies that dismantle the intervention by systematically manipulating the intervention's components will help identify which components of the intervention are necessary and sufficient to produce its distinct effects.

The intervention had a lasting effect on one measure of implicit race bias. Although many scholars have argued that implicit bias plays a pivotal role in the perpetuation of discrimination (Bargh, 1999; Devine, 1989; Fiske, 1998; Smedley et al., 2003), it will be important to demonstrate that reductions in implicit bias lead to reductions in discriminatory outcomes (e.g., interracial interaction quality, interview and hiring decisions, treatment in health settings).

In sum, this study presents the first intervention of its kind, one that, using a randomized controlled design, produces a reduction in implicit race bias that endures for at least two months. Our data provide evidence demonstrating the power of the conscious mind to intentionally deploy strategies to overcome implicit bias. As such, these findings raise the hope of solving a problem that has long vexed social scientists — how to reduce race-based discrimination. By empowering people to break the prejudice habit, this study takes an important step toward resolving the paradox of ongoing discrimination in a nation founded on the principle of equality.

Appendix A. Concern measure

Please indicate the degree to which you agree or disagree with each of the following statements using the scale below.

1	2	3	4	5	6	7	8	9	10
Strongly Disagree									strongly agree

1. I'm not personally concerned about discrimination against Blacks.
2. People need to stop focusing so much time and energy worrying about racial discrimination.
3. People make more fuss about discrimination against Blacks than is necessary.
4. I consider racial discrimination to be a serious social problem.

References

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A dynamic model of guilt: Implications for motivation and self-regulation in the context of prejudice. *Psychological Science*, *18*, 524–530, http://dx.doi.org/10.1111/j.1467-9280.2007.01933.x.

Bargh, J. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). New York: Guilford Press.

Beck, A. T., & Alford, B. A. (2009). *Depression: Causes and treatment* (2nd ed.). Baltimore, MD: University of Pennsylvania Press.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, *94*, 991–1013, http://dx.doi.org/10.1257/0002828042002561.

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

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*, 828–841, http://dx.doi.org/10.1037/0022-3514.81.5.828.

Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology*, *94*, 567–582, http://dx.doi.org/10.1037/a0014665.

Bradford, L. D., Newkirk, C., & Holden, K. B. (2009). Stigma and mental health in African Americans. In R. L. Braithwaite, S. E. Taylor, & H. M. Treadwell (Eds.), *Health issues in the black community*. San Francisco, CA: Jossey-Bass.

Brewer, M. (1988). A dual-process model of impression formation. In T. Srull, & R. Wyer (Eds.), *Advances in social cognition* (pp. 1–36). Hillsdale N.J.: Erlbaum Associates.

Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology*, *23*, 1933–1967, http://dx.doi.org/10.1111/j.1559-1816.1993.tb01074.x.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469–487, http://dx.doi.org/10.1037/0022-3514.89.4.469.

Cox, W. T. L., Abramson, L. Y., Devine, P. G., & Hollon, S. D. (2001). Stereotypes, Prejudice, and Depression: The Integrated Perspective. *Perspectives on Psychological Science*, *7*, http://dx.doi.org/10.1177/1745691612455204.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, *12*, 163–170, http://dx.doi.org/10.1111/1467-9280.00328.

Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience*, *16*, 1717–1729, http://dx.doi.org/10.1162/0898929042947919.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology*, *40*, 642–658, http://dx.doi.org/10.1016/j.jesp. 2004.02.003.

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, http://dx.doi.org/10.1037/0022-3514.81.5.800.

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

Devine, P. G., & Monteith, M. J. (1993). The role of discrepancy-associated affect in prejudice reduction. *Affect, cognition, and stereotyping: Interactive processes in group perception* (pp. 317–344). San Diego, CA: Academic Press.

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, http://dx.doi.org/10.1037/0022-3514.60.6.817.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*, 709–724, http://dx.doi.org/10.1037/0003-066X.49.8.709.

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, *25*, 1–54.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Fiske, S. (1998). Stereotyping, prejudice, and discrimination. In S. Fiske, D. Gilbert, & L. Gardner (Eds.), *The handbook of social psychology* (pp. 357–411). (4th ed.). Boston [etc.]: The McGraw-Hill.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology*, *Vol. 23*. (pp. 1–74)San Diego, CA: Academic Press.

Gaertner, S., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio, & S. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61–89). Orlando: Academic Press.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, *78*, 708–724, http://dx.doi.org/10.1037/0022-3514.78.4.708.

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

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

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, http://dx.doi.org/10.1037/0022-3514.85.2.197.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41, http://dx.doi.org/10.1037/a0015575.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality & Social Psychology Bulletin*, *31*, 1369–1385, http://dx.doi.org/10.1177/0146167205275613.

Howell, W. S. (1982). *The empathic communicator*. Prospect Heights, IL: Waveland.

Kaufman, D. M. (2003). ABC of learning and teaching in medicine: Applying educational theory in practice. *BMJ*, *326*, 213–216, http://dx.doi.org/10.1136/bmj.326.7382.213.

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

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, http://dx.doi.org/10.1006/jesp.2000.1470.

Mitchell, T. L., Haw, R. M., Pfeifer, J. F., & Meissner, C. A. (2005). Racial bias in mock juror decision-making: A meta-analytic review of defendant treatment. *Law and Human Behavior*, *29*, 621–637, http://dx.doi.org/10.1007/s10979-005-8122-9.

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, *65*, 469–485, http://dx.doi.org/10.1037/0022-3514.65.3.469.

Monteith, M. J., & Voils, C. I. (1998). Proneness to prejudiced responses: Toward understanding the authenticity of self-reported discrepancies. *Journal of Personality and Social Psychology*, *75*, 901–916, http://dx.doi.org/10.1037/0022-3514.75.4.901.

Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition*, *19*, 395–417, http://dx.doi.org/10.1521/soco.19.4.395.20759.

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*, 101–115, http://dx.doi.org/10.1037//1089-2699.6.1.101.

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, http://dx.doi.org/10.1146/annurev.psych.60.110707.163607.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*(2), 181–192, http://dx.doi.org/10.1037/0022-3514.81.2.181.

Pettigrew, T. F. (1998). Integroup contact theory. *Annual Review of Psychology*, *49*, 65–85, http://dx.doi.org/10.1146/annurev.psych.49.1.65.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*, 751–783, http://dx.doi.org/10.1037/0022-3514.90.5.751.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., et al. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, *12*, 729–738, http://dx.doi.org/10.1162/089892900562552.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*, 811–832, http://dx.doi.org/10.1037/0022-3514.75.3.811.

Plant, E. A., & Devine, P. G. (2009). The active control of prejudice: Unpacking the intentions guiding control efforts. *Journal of Personality and Social Psychology*, *96*, 640–652, http://dx.doi.org/10.1037/a0012960.

Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion*, *12*, 38–48, http://dx.doi.org/10.4278/0890-1171-12.1.38.

Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

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, http://dx.doi.org/10.1037/0022-3514.81.5.856.

Schuman, H., Steeh, C., Bobo, L., & Krysan, M. (1997). *Racial attitudes in America: Trends and interpretations* (Rev. ed.). Cambridge, MA: Harvard University Press.

Smedley, B., Stith, A., & Nelson, A. (Eds.). (2003). *Unequal treatment: Confronting racial and ethnic disparities in health care*. Washington D.C.: National Academy Press.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*, 108–131, http://dx.doi.org/10.1207/S15327957PSPR0402_01.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613–629, http://dx.doi.org/10.1037/0003-066X.52.6.613.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247, http://dx.doi.org/10.1207/s15327957pspr0803_1.

Van Breukelen, G. J. P. (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *Journal of Clinical Epidemiology*, *59*, 920–925, http://dx.doi.org/10.1016/j.jclinepi.2006.02.007.

Vontress, C., Woodland, C., & Epp, L. (2007). Cultural dysthymia: An unrecognized disorder among African Americans? *Journal of Multicultural Counseling and Development*, *35*, 130–141.