Social Psychological and
Personality Science
1-9
© The Author(s) 2017
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1948550617732817
journals.sagepub.com/home/spp
SAGE

Does Reducing Implicit Prejudice Increase Out-Group Identification? The Downstream Consequences of Evaluative Training on Associations Between the Self and Racial Categories

Curtis E. Phills[1], Kerry Kawakami[2], Danielle R. Krusemark[1], and John Nguyen[2]

Abstract

The present experiments were designed to investigate whether an intervention that targeted racial attitudes influenced not only prejudice but also self–Black associations. Because past research has demonstrated that people strive to build connections with favorable social categories, we predicted that positive evaluative training would increase identification with Blacks. Results from three studies provide evidence that practice in associating positive concepts with Blacks reduced implicit prejudice which in turn increased implicit self–Black associations. Notably, prejudice, in this case, had an intervening variable effect. Study 3 also investigated the impact of an alternative intervention that directly targeted self-associations rather than racial attitudes. Unlike evaluative training, associating the self with Blacks directly reduced both implicit prejudice and increased self–Black associations. These findings extend theorizing on the causal relationship between prejudice and out-group identification and provide important process information on how particular interventions reduce intergroup biases.

Keywords

prejudice reduction, identification, self-concept, evaluative conditioning

Often considered the most indispensable concept in social psychology (Allport, 1935; Briñol & Petty, 2012), attitudes provide valuable information on a variety of processes including how individuals visually process others (Young, Ratner, & Fazio, 2014), vote (Arcuri, Castelli, Galdi, Zogmaister, & Amadori, 2008), choose friends (Swann, Stein-Seroussi, & Giesler, 1992), and shop (Maison, Greenwald, & Bruin, 2004). Assessing attitudes, particularly implicit attitudes, is of particular importance in an intergroup context (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Although the relationship between implicit attitudes and behaviors is relatively small (Carlsson & Agerström, 2016; Greenwald et al., 2009; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015), implicit evaluations of racial/ethnic categories may be an important predictor of diverse spontaneous behaviors during cross-race interactions (Dovidio, Kawakami, & Gaertner, 2002; Kawakami, Amodio, & Hugenberg, 2017). The present research investigated another manner in which implicit attitudes may be important in an intergroup context: their potential causal relationship with out-group identification.

Along with implicit prejudice (i.e., negative associations with a particular social category), out-group identification is one of the most basic forms of bias. Perceiving out-group members as different and distinct from the self is a critical component of intergroup relations (Allport, 1954). Whether we believe that members of other groups have different personality traits, physical characteristics, cultural practices, goals, or values, a lack of correspondence between me and them can have a fundamental impact on processing in-group and out-group members (Kawakami et al., 2017; Van Bavel, Packer, & Cunningham, 2011).

Once a person is construed as a member of a social category, they are imbued with a wealth of categorical information (Fiske & Neuberg, 1990; Macrae & Bodenhausen, 2000) including not only group characteristics (stereotypes) but also evaluations (prejudice) and associations with the self (out-group

[1] University of North Florida, Jacksonville, FL, USA
[2] York University, Toronto, Ontario, Canada

Corresponding Author:
Curtis E. Phills, Department of Psychology, University of North Florida, 1 UNF Drive, Jacksonville, FL 32224, USA.
Email: curtis.phills@unf.edu

identification). Although these three constructs are considered to be distinct, empirical research related to the relationship between these types of biases is limited (Kawakami et al., 2017). This issue is particularly true for the link between prejudice and out-group identification. One potential strategy to investigate this relationship is to examine the impact of an intervention targeting one type of bias on the other.

Although recent research has highlighted a number of ways to reduce implicit prejudice (Brauer, Er-rafiy, Kawakami, & Phills, 2012; Lai et al., 2014; Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011; Phills, Santelli, Kawakami, Struthers, & Higgins, 2011), one particularly effective, direct method is evaluative conditioning—which pairs a target category with positive concepts to reduce negative attitudes (French, Franz, Phelan, & Blaine, 2013; Olson & Fazio, 2006). The primary goal of the present research was to investigate the impact of such an intervention not only on prejudice but also on identification. In the current studies, prejudice was measured with an attitude implicit association test (IAT) and was related to the speed of associating positivity compared to negativity more with one social category (Whites) than another (Blacks). Identification bias, alternatively, was measured with an identity IAT and was related to the speed of associating the self compared to others more with one social category (Whites) than another (Blacks).

Previous theorizing and research suggest that attitudes may be causally related to identification; to maintain and enhance their self-image, people surround themselves with positive possessions and people (Cialdini & Richardson, 1980; Kelley & Thibaut, 1978). As indicated by research on basking in reflected glory (Cialdini et al., 1976; Cialdini & De Nicholas, 1989), people not only prefer to associate with more favorable individuals but also successful in-groups (e.g., a winning college football team). Because people believe connections to favorable in-groups and distance from unfavorable in-groups makes them look good (Snyder, Lassegard, & Ford, 1986; Spears, Doosje, & Ellemers, 1997), we predicted that interventions that directly target and improve racial attitudes would increase out-group identification.

Notably, research has also investigated strategies targeting out-group identification (Galinsky, Wang, & Ku, 2008; Greenwald, Pickrell, & Farnham, 2002). For example, studies have demonstrated that self-out-group associations increased via training to conceptually approach out-group members (i.e., pulling out-group members toward the self in a joystick paradigm; Phills, Kawakami, et al., 2011) and taking the perspective of out-group members; Todd & Burgmer, 2013) not only to enhance out-group identification but also to reduce prejudice. A secondary goal of the present research was to investigate the impact of a novel more direct strategy to increase identification (i.e., training in associating self and others with Blacks and Whites) on both identification and racial attitudes.

One reason why interventions focusing on identification may result in more positive racial attitudes is that people tend to associate positive, and not negative, evaluations with the self (Bosson, Swann, & Pennebaker, 2000; Ye & Gawronski, 2016). Because research has demonstrated that increasing associations between the self- and out-groups increases the transfer of associations with the self to racial categories such as Blacks (Gawronski, Bodenhausen, & Becker, 2007; Phills, Kawakami, et al., 2011), we predicted that strategies that directly target and improve out-group identification would also decrease prejudice.

In summary, three experiments investigated the causal relationship between implicit prejudice and out-group identification. In particular, Studies 1 and 2 examined whether evaluative training in associating positive but not negative concepts with Blacks would reduce implicit prejudice, which in turn would increase Black–self associations. Study 3 also included an intervention that directly targeted Black–self associations to investigate the bidirectionality of the relationship between attitudes and identification.

Together, these experiments have the potential to provide new evidence for a close and causal relationship between two important intergroup biases, as well as to inform us about processes related to bias reduction. While previous research has often investigated the possibility that interventions aimed at either decreasing prejudice or enhancing identification may positively influence the targeted bias, the present studies explore the broader consequences of these methods.

Study 1

Method

Participants and procedure. Because of the dearth of research on the impact of changes in implicit prejudice on out-group identification, we did not have a reliable estimate of effect size. We therefore initially aimed to recruit approximately 50 participants in each condition (Simmons, Nelson, & Simonsohn, 2013). Although 107 non-Black Canadian undergraduates participated, 18 students were excluded for not completing both IATs or making more than 40% errors on the IATs. The final sample included 89 participants (77 female; $M_{\text{age}} = 21.45$, $SD = 10.99$; 21 East Asian, 3 Hispanic, 10 Middle Eastern, 37 South Asian, and 18 White). Participants were randomly assigned to either a Black positive or Black negative training condition before being presented with two IATs related to attitudes and identity in a counterbalanced order.

Black Evaluative Training Task. Participants were presented in this task with a series of photographs of single faces in the center of a computer screen. One positive and one negative word were positioned an equal number of times across trials below the image on the left or right side. In contrast to most evaluative conditioning paradigms (Olson & Fazio, 2006), participants were instructed to actively select either positive or negative concepts depending on the target group. Specifically, participants in the Black positive condition were required to select a positive word when presented with a Black face and a negative word when presented with a White face. Participants in the Black negative condition were given the opposite instructions. The stimuli remained on the screen until participants responded. If the response was correct, a blank screen appeared for 1,000 ms before the next trial. If the response was incorrect, a blank screen appeared for 100 ms, followed by a red “X” in the center of the screen for 800 ms, and a blank screen for 100 ms.

Participants completed six blocks of 80 trials (480 trials). The stimuli included 48 faces (24 Black, 24 White) and 20 positive (caress, freedom, love, peace, cheer, loyal, pleasure, gentle, honest, vacation, lucky, rainbow, gift, honor, miracle, sunrise, family, happy, laughter, and paradise) and 20 negative (abuse, crash, filth, sickness, accident, death, grief, poison, stink, disaster, hatred, pollute, tragedy, bomb, divorce, ugly, cancer, evil, rotten, and vomit) words unrelated to Black or White stereotypes.

Attitude IAT. To assess implicit prejudice toward Blacks and Whites (Greenwald, McGhee, & Schwartz, 1998; Nosek et al., 2007), participants were instructed to categorize six photographs of Blacks (three male and three female) and Whites (three male and three female), as well as six positive (beautiful, marvelous, wonderful, glorious, lovely, and superb) and six negative (disgust, pain, terrible, horrible, hate, and awful) words not included in the training.

Following standard IAT procedures, participants completed five blocks (three practice). Incongruent critical blocks required participants to use one key to categorize Blacks and positive words and another key to categorize Whites and negative words. Congruent critical blocks required participants to use one key to categorize Whites and positive words and another key to categorize Blacks and negative words. The order of the incongruent and congruent blocks was counterbalanced. Each stimulus was presented 3 times during the critical blocks (72 trials). Procedures related to incorrect responding were the same as in the evaluative training.

Identity IAT. To assess self-out-group associations (Greenwald & Farnham, 2000), participants were instructed to categorize the photographs of Blacks and Whites included in the attitude IAT and four words related to the self (I, me, mine, self, and they) and others (they, them, their, and others). Research has shown that the number of stimuli does not significantly influence the magnitude of IAT effect size (Nosek, Greenwald, & Banaji, 2005).

In this IAT, incongruent critical blocks required participants to use the same key to categorize Blacks and self-related words and another key to categorize Whites and other-related words. Congruent critical blocks required participants to use the same key to categorize Whites and self-related words and another key to categorize Blacks and other-related words. Each stimulus was presented 3 times during the critical blocks (60 trials). The number of blocks, their order, and response feedback were identical to the attitude IAT.

Results

IAT scores in all experiments were calculated according to a standard algorithm (Greenwald, Nosek, & Banaji, 2003) with higher scores representing more positive attitudes and greater identification with Whites over Blacks. In this study, attitude, $t(88) = 4.85$, $p \leq .001$, and identity, $t(88) = 1.88$, $p = .063$, IAT scores significantly and marginally differed from 0, suggesting that participants associated less positive concepts and identified less with Blacks. Despite attitudes and identification both being measured with an IAT, because they are distinct constructs (Cohen, 2001), we elected to conduct separate 2 (training: Black positive vs. Black negative) × 2 (IAT order: attitude vs. identity IAT first) analyses of variance (ANOVAs) on each IAT score.

Attitude IAT. The ANOVA on attitude IAT scores demonstrated that participants trained to associate positive ($D = .05$, $SD = 0.42$) compared to negative ($D = .30$, $SD = 0.32$) concepts with Blacks had lower implicit prejudice, $F(1, 85) = 7.11$, $p = .009$, $\eta_{\mathrm{p}}^2 = .08$. The main effects of IAT order, $F(1, 85) = 1.22$, $p = .272$, $\eta_{\mathrm{p}}^2 = .01$, and the interaction, $F(1, 85) = 0.16$, $p = .695$, $\eta_{\mathrm{p}}^2 < .01$, were not significant.

Identity IAT. The ANOVA on identity IAT scores showed no significant difference in Black–self associations between participants trained to associate positive ($D = .03$, $SD = 0.28$) or negative ($D = .07$, $SD = 0.26$) concepts with Blacks, $F(1, 85) = 0.11$, $p = .743$, $\eta_p^2 < .01$. Although a main effect of IAT order demonstrated that participants who completed the identity IAT ($D = -.003$, $SD = 0.24$) rather than the attitude IAT ($D = .14$, $SD = 29$) first had stronger Black–self associations, $F(1, 85) = 7.30$, $p = .008$, $\eta_p^2 = .08$, the interaction was not significant, $F(1, 85) = 1.63$, $p = .205$, $\eta_p^2 = .02$.

Correlational and mediational analyses. Analyses examining correlations between attitude and identity IAT scores indicated that less positive attitudes were related to less identification with Blacks, $r(89) = .39$, $p < .001$. Correlation magnitude did not differ between participants trained to associate positive, $r(37) = .32$, $p = .052$, and negative, $r(52) = .46$, $p = .001$, concepts with Blacks, $Z = .73$, $p = .465$.

We used PROCESS Model 4 (Hayes, 2013) to test whether attitudes played an indirect role in the relationship between Black evaluative training and out-group identification. Because the independent variable did not significantly affect the primary dependent variable, the present analyses investigated an intervening variable effect in which training indirectly influenced out-group identification through implicit prejudice (Pek & Hoyle, 2016). Thus, we did not expect implicit prejudice to explain the effect of evaluative training on out-group identification (because there was none), but rather we examined whether evaluative training reduced attitude IAT scores which in turn reduced identity IAT scores.

To assess the significance of this indirect effect (Figure 1, Panel A), we used 5,000 bootstrapped resamples to generate a 95% confidence interval (CI) = [.02, .14]. Consistent with the proposed hypothesis, this interval did not include 0 providing initial evidence that training in associating positive concepts with Blacks reduced implicit prejudice and that these more

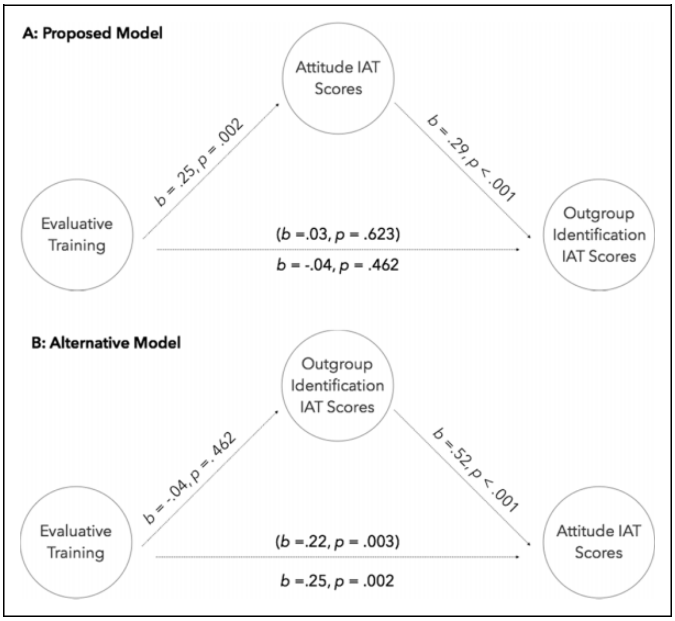

Figure 1. Unstandardized regression coefficients in Study 1 for the relationship between evaluation training, attitude IAT scores, and identity IAT scores. Panel A depicts the proposed model with attitude IAT scores as the intervening variable and Panel B depicts the alternative model with identity IAT scores as the intervening variable.

positive attitudes increased identification with Blacks. A separate model with IAT order as a moderator (PROCESS Model 59; Hayes, 2013) did not significantly differ depending on IAT order, with the CI around the indirect effect including 0, 95% CI [−.34, .04]. Moreover, an alternative model with identity IAT scores as the mediator and attitude IAT scores as the dependent variable (Figure 1, Panel B) was not significant, 95% CI [−.05, .07].

Study 2

Method

Participants and procedure. The primary goal of Study 2 was to replicate the findings of Study 1. Based on the bootstrapping analysis of Study 1 (standardized coefficients of $\alpha = -.32$ and $\beta = .41$), to achieve 80% power, 115 participants were needed (Fritz & MacKinnon, 2007). However, we recruited a larger sample to offset fewer trials in the training task and to explore potential moderating effects of individual differences in self-concept constructs. Although, 276 non-Black U.S. MTurk workers were recruited, 37 were excluded because they failed to complete both IATs or because they exceeded 40% errors during those tasks. The final sample included 239 (139 female; $M_{\text{age}} = 41.44$, $SD = 12.27$; 4 First Nation, 10 East Asian, 9 Hispanic, 22 South Asian, and 194 White) participants.

Although participants were again randomly assigned to either the Black positive or Black negative training conditions, the procedure differed from Study 1 in four ways. First, the training task consisted of 5 blocks of 48 trials (240 total trials). Second, a different set of photographs of Black and White targets were included in the training task (Westfall, Judd, & Kenny, 2015). Third, before completing the training, participants completed several self-concept individual differences measures. Exploratory analyses indicated that on both the attitude and identity IATs, the interaction between evaluative training and each of the following measures was not significant: self-concept clarity (Campbell et al., 1996), ps = .146 and .097, sense of self (Flury & Ickes, 2007), ps = .147 and .380, perspective taking (Davis, 1983), ps = .394 and .941, need for affiliation (Hill, 1987), ps = .910 and .137, self-monitoring (Lennox & Wolfe, 1984), ps = .617 and .166, and contingent self-worth (Crocker, Luhtanen, Cooper, & Bouvrette, 2003), ps = .978 and .257. Fourth, although the order of the IATs was counterbalanced across conditions, the order of the blocks within each IAT was held constant.

Results

Initial analyses comparing IAT scores to 0 demonstrated bias in both attitude scores, $t(238) = 8.72$, $p < .001$, and identity scores, $t(238) = 6.39$, $p < .001$, suggesting that participants associated less positive concepts and identified less with Blacks. Again, a 2 (training: Black positive vs. Black negative) × 2 (IAT order: attitude vs. identity IAT first) ANOVA was conducted on each IAT score.

Attitude IAT. The ANOVA on attitude IAT scores demonstrated that training to associate positive ($D = .16$, $SD = 0.38$) compared to negative ($D = .28$, $SD = 0.41$) concepts with Blacks resulted in lower implicit prejudice, $F(1, 235) = 6.011$, $p = .015$, $\eta_p^2 = .03$. The main effects of IAT order, $F(1, 235) = 1.71$, $p = .192$, $\eta_p^2 = .01$, and the interaction, $F(1, 235) = 2.03$, $p = .156$, $\eta_p^2 = .01$, were not significant.

Identity IAT. The ANOVA on identity IAT scores demonstrated no difference in Black–self associations between participants trained to associate positive ($D = .12$, $SD = 0.37$) and negative ($D = .18$, $SD = 0.36$) concepts with Blacks, $F(1, 235) = 1.60$, $p = .207$, $\eta_p^2 = .01$. In addition, IAT order, $F(1, 235) = 0.15$, $p = .702$, $\eta_p^2 < .01$, and the interaction, $F(1, 235) = 1.639$, $p = .202$, $\eta_p^2 = .01$, were not significant.[1]

Correlational and mediational analyses. Analyses examining correlations between attitude and identity IAT scores suggest that less positive attitudes were related to less identification with Blacks, $r(239) = .26$, $p < .001$. The correlation magnitude did not differ between Black positive, $r(114) = .24$, $p = .010$, and Black negative, $r(125) = .24$, $p = .008$, training conditions, $Z = .03$ $p = .976$.

To investigate the indirect effects of Black evaluative training on out-group identification via implicit prejudice (Figure 2, Panel A), 5,000 resamples were used to generate a 95% CI [.01, .06]. This significant indirect effect provided further evidence for the intervening variable role of implicit prejudice by suggesting that practice in associating positive concepts with Blacks reduces implicit prejudice, which in turn increases

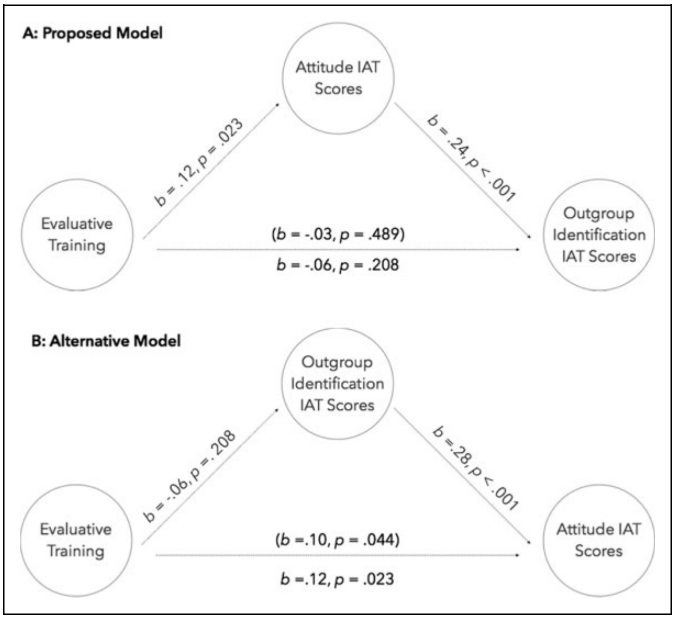

Figure 2. Unstandardized regression coefficients in Study 2 for the relationship between evaluation training, attitude IAT scores, and identity IAT scores. Panel A depicts the proposed model with attitude IAT scores as the intervening variable and Panel B depicts the alternative model with identity IAT scores as the intervening variable.

self–Black associations. A separate model (PROCESS Model 59) with IAT order as a moderator did not differ depending on IAT order, 95% CI [−.03, .11]. A test of an alternative model which included identity IAT scores as the intervening variable and attitude IAT scores as the dependent variable (Figure 2, Panel B), 95% CI [−.05, .02], was also not significant.

Study 3

Method

Participants and procedure. Study 3 sought to replicate the effects of evaluation training as well as investigate the impact of an intervention that directly targeted self-associations rather than racial attitudes. Specifically, Study 3 explored whether identity training results in the same pattern of changes in bias as evaluative training with a direct effect on attitudes and an indirect effect on identification, the reverse pattern with a direct effect on identity and an indirect on attitudes, or perhaps because of the importance of the self-concept, a direct effect on both identity and attitudes.

Because Study 3 recruited from the same population of MTurk workers as Study 2, we conducted a power analysis based on the experiment's standardized regression coefficients ($\alpha = -.15$ and $\beta = .25$). To achieve 80% power, 754 participants were needed (Fritz & Mackinnon, 2007). Although 736 participants were recruited, 17 did not finish the experiment and 44 were excluded for exceeding 40% IAT errors. The remaining 675 (408 female; 10 First Nation, 46 Asian, 33 Hispanic, and 586 White) participants were randomly assigned to complete a training either targeting attitudes or identity in a 2 (type of training: Black evaluative vs. Black–self) × 2 (training goal: reduce bias vs. maintain bias) between-groups design.

Although both types of training consisted of 4 blocks of 48 trials (192 trials), half of the participants completed a training related to associating Blacks with evaluative concepts and half completed a training related to associating the self with Blacks. While the evaluative training was similar to the task used in Studies 1 and 2, in the later training, participants either were instructed to select "me" when presented with a photograph of a Black person and "not me" when presented with a photograph of a White person (Black me) or were given the opposite instructions (Black not me). Whereas the goal of Black positive and Black me training was to reduce bias, the goal of Black negative and Black not me training was to maintain bias. Following the training, participants completed the same attitude and identity IATs from Study 2 in a counterbalanced order.

Results

Initial analyses demonstrated that both attitude, $t(674) = 13.12$, $p < .001$, and identity, $t(674) = 5.87$, $p < .001$, IAT scores differed from 0, suggesting implicit prejudice and less identification with Blacks. A 2 (Type of training: Black evaluative vs. Black–self) × 2 (Training goal: reduce bias vs. maintain bias) × 2 (IAT order: attitude vs. identity IAT first) ANOVA was conducted on each IAT score.

Attitude IAT. The ANOVA on attitude IAT scores demonstrated a main effect of training goal, $F(1, 667) = 11.53, p = .001, \eta_p^2 = .02$. Participants in the bias reduction ($D = .15$, $SD = 0.39$) compared to bias maintenance ($D = .26$, $SD = 0.44$) conditions showed less implicit prejudice. The main effects of type of training, $F(1, 667) = 0.45$, $p = .503$, $\eta_p^2 < .01$, IAT order, $F(1, 667) = 2.19$, $p = .139$, $\eta_p^2 < .01$, and all interactions, $ps > .110$, were not significant.

Identity IAT. The ANOVA on identity IAT scores demonstrated marginal or significant main effects of training goal, $F(1, 667) = 2.97$, $p = .086$, $\eta_p^2 < .01$, type of training, $F(1, 667) = 3.58$, $p = .058$, $\eta_p^2 = .01$, and IAT order, $F(1, 667) = 20.81$, $p < .001$, $\eta_p^2 = .03$. Participants in bias reduction ($D = .06$, $SD = 0.34$) compared to bias maintenance ($D = .10$, $SD = 0.38$) conditions and participants who completed the Black–self ($D = .06$, $SD = 0.35$) compared to the Black evaluative ($D = .10$, $SD = 0.37$) training had somewhat higher out-group identification. Participants who completed the attitude ($D = .02$, $SD = 0.37$) compared to the identity ($D = .14$, $SD = 0.34$) IAT first had higher out-group identification.

Importantly, the Type of Training × Training Goal interaction was also significant, $F(1, 667) = 5.73$, $p = .017$, $\eta_p^2 = .01$. As in Studies 1 and 2, training to associate positive ($D = .11$, $SD = 0.34$) rather than negative ($D = .09$, $SD = 0.41$) concepts with Blacks did not influence identification IAT scores, $F(1, 667) = 0.23, p = .631, \eta_p^2 < .01$. However, training to associate Blacks with me ($D = .01$, $SD = 0.34$) compared to not me ($D = .11$, $SD = 0.35$) increased out-group identification,

$F(1, 667) = 8.28$, $p = .004$, $\eta_p^2 = .01$. No other interactions were significant, *p*s > .161.[2]

Correlation and mediation analyses. Correlations between attitude and identity IAT scores, $r(675) = .10$, $p = .008$, indicated that less positive attitudes were related to less identification with Blacks, in general, and for participants who received Black evaluation training, $r(342) = .14$, $p = .011$. The magnitude of this coefficient did not differ between Studies 2 and 3, $Z = 1.48$, $p = .139$. Moreover, the correlation magnitude did not differ for participants in the Black positive, $r(177) = .18$, $p = .014$, and Black negative, $r(165) = .11$, $p = .177$, training conditions, $Z = .76$, $p = .447$. Although the correlation was not significant for participants who completed the Black–self training, in general, $r(333) = .06$, $p = .309$, this correlation was significant for participants in the Black me training condition $r(158) = .17$, $p = .033$, and different than the Black not me condition, $r(175) = -.06$, $p = .400$, $Z = 2.13$, $p = .033$.

We calculated separate mediation models for participants that completed evaluative and self-associations training using PROCESS Model 4 (Hayes, 2013).

Black evaluative training. The 95% CI [−.03, −.002] created with 5,000 resamples related to the indirect effect of Black evaluative training on out-group identification through attitudes (Figure 3, Panel A) was significant and provided further evidence that implicit prejudice plays an intervening variable role. A separate model (PROCESS Model 59) with IAT order as a moderator did not differ depending on IAT order, 95% CI [−.01, .06]. Furthermore, a test of an alternative model with identity IAT scores as the mediator and attitude IAT as the dependent variable (Figure 3, Panel B), 95% CI [−.02, .01], was not significant, suggesting no mediation.

Black–self training. Although training in Black me compared to Black not me associations had an impact on both attitude and identity IAT scores, attitude IAT scores did not mediate the impact of Black–self training on identity IAT scores (Figure 3, Panel C), 95% CI [−.01, .02], and identity IAT scores did not mediate the impact of Black–self training on attitude IAT scores (Figure 3, Panel D), 95% CI [−.01, .02]. Separate models (PROCESS Model 59) with IAT order as a moderator did not differ depending on IAT order when attitude IAT scores were the mediator, 95% CI [−.02, .03], or when identity IAT scores were the mediator, 95% CI [−.04, .03].

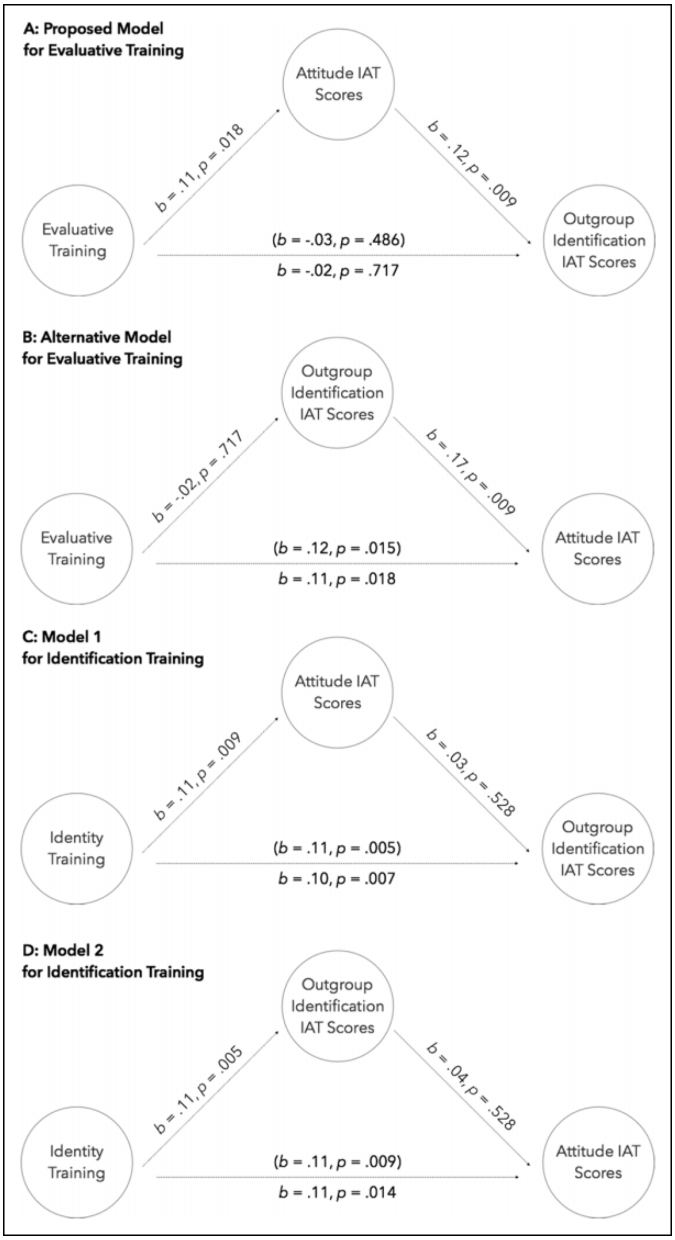

Figure 3. Unstandardized regression coefficients in Study 3 for the relationship between training, attitude IAT scores, and identity IAT scores. Panel A depicts the proposed model with attitude IAT scores as the intervening variable and evaluative training. Panel B depicts the alternative model with identity IAT scores as the intervening variable and evaluative training. Panel C depicts Model 1 with attitude IAT scores as the mediating variable and identity training. Panel D depicts Model 2 with identity IAT scores as the mediating variable and identity training.

General Discussion

Three experiments with diverse samples provided consistent evidence that decreasing implicit prejudice via evaluative training increased out-group identification indirectly. Combining the data from all studies, the effect size of evaluative training on implicit prejudice was between small and medium (*d* = .29) and similar to earlier reports (*d* = .21, Lai et al., 2014). In contrast, although evaluative training did not significantly impact out-group identification in any of the experiments or when the data were combined, $t(491) = 0.17$, $p = .866$, it did have an indirect effect through implicit prejudice in each experiment and when the data were combined, 95% CI [.01, .05].

Notably, in Study 3, we also investigated whether interventions targeting out-group identification rather than attitudes would work in a conceptually similar way. The answer is no.

Although training in associating Blacks with me significantly impacted both implicit prejudice and out-group identification, these variables did not mediate one another.

One possible caveat to interpreting the present findings may be related to method-specific sources of variances in the IAT. In particular, it is possible that IAT scores may predict one another due to individual differences in cognitive skills and task-switching abilities rather than their content. However, this interpretation is countered by the findings that evaluative training had a direct impact on attitude but not identity IAT scores in all three experiments and that correlations between IAT scores were not consistently significant in Study 3. Thus, despite having some structural overlap, the IATs in the present context are not only conceptually distinct but influenced in unique ways by different interventions. Nonetheless, we recommend that future research includes alternative ways of operationalizing implicit prejudice and out-group identification.

Although the present research was not designed to test the Balanced Identity Theory (BIT; Cvencek, Greenwald, & Meltzoff, 2012), it provides new causal evidence for the close causal relationship between implicit identities and attitudes in an intergroup context. While the BIT suggests that the three legs of a triad between identities, attitudes, and self-esteem organize to maintain affective-cognitive consistency and that the interrelationships between these constructs constrain each other, the present research did not examine associations between the self and positivity. Although research by Dunham (2013) using a minimal group paradigm provides evidence that targeting attitudes, self-esteem, and identification can create a balanced model, the current experiments, alternatively, investigated the potential bidirectionality of the relationship between attitudes and identities.

Together the current studies demonstrate how a single intervention can impact diverse intergroup biases and the process through which this occurs. Because interventions are typically used to ameliorate a specific form of differential responding to Blacks and Whites, their capacity to change other biases and possibly even intergroup behavior is often underappreciated. For example, researchers have limited their investigation of the effects of evaluative conditioning for the most part to attitudes, ignoring its potential to decrease a host of other biases. Because certain types of bias may be distinctly associated with particular behaviors, it is critical to discover new ways to change them (Dovidio et al., 2002; Kawakami, Phills, Steele, & Dovidio, 2007). Although these questions highlight the fact that we are just in the early stages of understanding different permutations of racial bias, how they can be modified, and how they relate, the present findings provide hope that one intervention can potentially have a range of implications for developing positive intergroup relations.

Notes

1. Although there were too few students in Study 1 to examine these effects with only White participants, when analyses were limited to this group in Study 2, evaluative training continued to influence attitude IAT scores, $F(1, 190) = 11.24, p = .001, \eta_p^2 = .06$, but not identity IAT scores, $F(1, 190) = 1.387, p = .240, \eta_p^2 = .01$.
2. When only White participants were included in the analyses, the pattern of findings remains the same. Black evaluative training influenced attitude IAT scores, $F(1, 578) = 3.91, p = .048, \eta_p^2 = .01$, but not identity IAT scores, $F(1, 578) = 0.50, p = .479, \eta_p^2 < .01$. Alternatively, Black–self training influenced both attitude IAT scores, $F(1, 578) = 5.86, p = .016, \eta_p^2 = .01$, and identity IAT scores, $F(1, 578) = 4.27, p = .039, \eta_p^2 = .01$.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Allport, G. W. (1935). Attitudes. In C. M. Murchison (Ed.), *Handbook of Social Psychology*. Winchester, MA: Clark University Press.

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

Arcuri, L., Castelli, L., Galdi, S., Zogmaister, C., & Amadori, A. (2008). Predicting the vote: Implicit attitudes as predictors of the future behavior of decided and undecided voters. *Political Psychology*, *29*, 369–387. doi:10.1111/j.1467-9221.2008.00635.x

Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited?*Journal of Personality and Social Psychology*, *79*, 631–643. doi:10.1037/0022-3514.79.4.631

Brauer, M., Er-rafiy, A., Kawakami, K., & Phills, C. E. (2012). Describing a group in positive terms reduces prejudice less effectively than describing it in positive and negative terms. *Journal of Experimental Social Psychology*, *48*, 757–761. doi:10.1016/j.jesp.2011.11.002

Briñol, P., & Petty, R. E. (2012). The history of attitudes and persuasion research. In A. Kruglanski & W. Stroebe (Eds.), *Handbook of the History of Social Psychology*. New York, NY: Psychology Press.

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology*, *70*, 141–156.

Carlsson, R., & Agerström, J. (2016). A closer look at the discrimination outcomes in the IAT literature. *Scandinavian Journal of Psychology*, *57*, 278–287. doi:10.1111/sjop.12288

Cialdini, R. B., Borden, R. J., Thorne, A., Walker, M. R., Freeman, S., & Sloan, L. R. (1976). Basking in reflected glory: Three (football) field studies. *Journal of Personality and Social Psychology*, *34*, 366–375. doi:10.1037/0022-3514.34.3.366

Cialdini, R. B., & de Nicholas, M. E. (1989). Self-presentation by association. *Journal of Personality and Social Psychology*, *57*, 626–631. doi:10.1037/0022-3514.57.4.626

Cialdini, R. B., & Richardson, K. D. (1980). Two indirect tactics of image management: Basking and blasting. *Journal of Personality* *and Social Psychology*, *39*, 406–415. doi:10.1037/0022-3514.39.3.406

Cohen, B. H. (2001). *Explaining psychological statistics* (2nd ed.). New York, NY: John Wiley.

Crocker, J., Luhtanen, R. K., Cooper, M. L., & Bouvrette, A. (2003). Contingencies of self-worth in college students: Theory and measurement. *Journal of Personality and Social Psychology*, *85*, 894–908. doi:10.1037/0022-3514.85.5.894

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2012). Balanced Identity Theory. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 157–177). Guilford, NY: Guilford Press.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113–126. doi:10.1037/0022-3514.44.1.113

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*, 62–68. doi:10.1037/0022-3514.82.1.62

Dunham, Y. (2013). Balanced identity in the minimal groups paradigm. *PLoS One*, *8*. doi:10.1371/journal.pone.0084205

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. *Advances in Experimental Social Psychology*, *23*, 1–74. doi:10.1016/S0065-2601(08)60317-2

Flury, J. M., & Ickes, W. (2007). Having a weak versus strong sense of self: The sense of self scale (SOSS). *Self and Identity*, *6*, 281–303. doi:10.1080/15298860601033208

French, A. R., Franz, T. M., Phelan, L. L., & Blaine, B. E. (2013). Reducing Muslim/Arab stereotypes through evaluative conditioning. *The Journal of Social Psychology*, *153*, 6–9. doi:10.1080/00224545.2012.706242

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*, 233–239. doi:10.1111/j.1467-9280.2007.01882.x

Galinsky, A. D., Wang, C. S., & Ku, G. (2008). Perspective-takers behave more stereotypically. *Journal of Personality and Social Psychology*, *95*, 404–419. doi:10.1037/0022-3514.95.2.404

Gawronski, B., Bodenhausen, G. V., & Becker, A. P. (2007). I like it, because I like myself: Associative self-anchoring and post-decisional change of implicit evaluations. *Journal of Experimental Social Psychology*, *43*, 221–232. doi:10.1016/j.jesp.2006.04.001

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*, 1022–1038. doi:10.1037//0022-3514.79.6.I022

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*, 481–481. doi:10.1037/h0087889

Greenwald, A. G., Pickrell, J. E., & Farnham, S. D. (2002). Implicit partisanship: Taking sides for no reason. *Journal of Personality and Social Psychology*, *83*, 367–79. doi:10.1037/0022-3514.83.2.367

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. doi:10.1037/a0015575

Greenwald, A., McGhee, D., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480. doi:10.1027/1015-5759.24.4.210

Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Guilford Press. doi:978-1-60918-230-4

Hill, C. A. (1987). Affiliation motivation: People who need people-but in different ways. *Journal of Personality and Social Psychology*, *52*, 1008–1018. doi:10.1037/0022-3514.52.5.1008

Kawakami, K., Amodio, D., & Hugenberg, K. (2017). Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of social categorization. *Advances in Experimental Social Psychology*, *55*, 1–80.

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology*, *92*, 957–71. doi:10.1037/0022-3514.92.6.957

Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal relations: A theory of interdependence*. New York, NY: John Wiley.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. L., Joy-gaba, J. A., . . . Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*, 1765–1785. doi:10.1037/a0036260

Lennox, R. D., & Wolfe, R. N. (1984). Revision of the self-monitoring scale. *Journal of Personality and Social Psychology*, *46*, 1349–1364. doi:10.1037/0022-3514.46.6.1349

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120. doi:10.1146/annurev.psych.51.1.93

Maison, D., Greenwald, A. G., & Bruin, R. H. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology*, *14*, 405–415. doi:10.1207/s15327663jcp1404_9

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

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

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

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of*

Personality and Social Psychology, *108*, 562–571. doi:10.1037/pspa0000023

Pek, J., & Hoyle, R. H. (2016). On the (in)validity of tests of simple mediation: Threats and solutions. *Social and Personality Psychology Compass*, *10*, 150–163.

Phills, C. E., Kawakami, K., Tabi, E., Nadolny, D., & Inzlicht, M. (2011). Mind the gap: Increasing associations between the self and blacks with approach behaviors. *Journal of Personality and Social Psychology*, *100*, 197–210. doi:10.1037/a0022159

Phills, C. E., Santelli, A. G., Kawakami, K., Struthers, C. W., & Higgins, E. T. (2011). Reducing implicit prejudice: Matching approach/avoidance strategies to contextual valence and regulatory focus. *Journal of Experimental Social Psychology*, *47*. doi:10.1016/j.jesp.2011.03.013

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2013). Life after p-hacking. *SSRN Electronic Journal*, *41*, 775. doi:10.2139/ssrn.2205186

Snyder, C. R., Lassegard, M., & Ford, C. E. (1986). Distancing after group success and failure: Basking in reflected glory and cutting off reflected failure. *Journal of Personality and Social Psychology*, *51*, 382–388. doi:10.1037/0022-3514.51.2.382

Spears, R., Doosje, B., & Ellemers, N. (1997). Self-stereotyping in the face of threats to group status and distinctiveness: The role of group identification. *Personality and Social Psychology Bulletin*, *23*, 538–553. doi:10.1177/0146167297235009

Swann, W. B., Stein-Seroussi, A, & Giesler, R. B. (1992). Why people self-verify. *Journal of Personality and Social Psychology*, *62*, 392–401. doi:10.1037/0022-3514.62.3.392

Todd, A. R., & Burgmer, P. (2013). Perspective taking and automatic intergroup evaluation change: Testing an associative self-anchoring account. *Journal of Personality and Social Psychology*, *104*, 786–802. doi:10.1037/a0031999

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2011). Modulation of the fusiform face area following minimal exposure to motivationally relevant faces: Evidence of in-group enhancement (not out-group disregard). *Journal of Cognitive Neuroscience*, *23*, 3343–3354. doi:10.1162/jocn_a_00016

Westfall, J., Judd, C. M., & Kenny, D. (2015). Replicating studies in which samples of participants respond to samples of stimuli. *Perspectives on Psychological Science*, *10*, 390–399. doi:10.1177/1745691614564879

Ye, Y., & Gawronski, B. (2016). When possessions become part of the self: Ownership and implicit self-object linking. *Journal of Experimental Social Psychology*, *64*, 72–87. doi:10.1016/j.jesp.2016.01.012

Young, A. I., Ratner, K. G., & Fazio, R. H. (2014). Political attitudes bias the mental representation of a presidential candidate's face. *Psychological Science*, *25*, 503–510. doi:10.1177/0956797613510717

Author Biographies

Curtis E. Phills is an assistant professor of psychology at the University of North Florida in Jacksonville, FL.

Kerry Kawakami is a full professor of psychology at York University in Toronto, Ontario, Canada.

Danielle R. Krusemark was an undergraduate honors thesis student at the University of North Florida in Jacksonville, FL. She is now a graduate student at Florida State University.

John Nguyen was an undergraduate honors thesis student at York University in Toronto, Ontario, Canada.

Handling Editor: Kate Ratliff