Article

The Bidirectional Causal Relation Between Implicit Stereotypes and Implicit Prejudice

Personality and Social
Psychology Bulletin
2020, Vol. 46(9) 1318–1330
© 2020 by the Society for Personality
and Social Psychology, Inc
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0146167219899234
journals.sagepub.com/home/pspb
SAGE

Curtis E. Phills[1]*, Adam Hahn[2]*, and Bertram Gawronski[3]

Abstract

Although stereotypes and prejudice are commonly regarded as conceptually distinct but related constructs, previous research remains silent on the processes underlying their relation. Applying the balance-congruity principle to the concepts (a) group, (b) valence, and (c) attribute, we argue that the valence of attributes contained in a group-stereotype shapes evaluations of the group, while prejudice toward a group influences which attributes are stereotypically associated with the group. Using fictitious (Experiments 1 and 3) and real (Experiments 2 and 4) groups, the current studies demonstrate that (a) experimentally induced changes in the valence of semantic attributes associated with a group (stereotypes) influence implicit prejudice toward that group (Experiments 1 and 2), and (b) experimentally induced changes in the valence of a group (prejudice) influence implicit stereotyping of that group (Experiments 3 and 4). These findings demonstrate a bidirectional causal relation between prejudice and stereotypes.

Keywords

balance-congruity principle, evaluative conditioning, implicit social cognition, prejudice, stereotypes

Received May 4, 2019; revision accepted December 13, 2019

Extant theories in social psychology suggest that implicit evaluations are rooted in mental associations of social groups with positive or negative valence, whereas implicit stereotypes reflect mental associations of social groups with semantic attributes (Amodio & Devine, 2006; Greenwald et al., 2002). For example, negative implicit evaluations of African Americans may be rooted in mental associations between African Americans and negative valence, whereas implicit stereotypes may be rooted in mental associations of African Americans with stereotypical traits (e.g., athletic, musical). Although the two kinds of associations are commonly treated as distinct in terms of their contents (i.e., evaluative vs. semantic), there is evidence that implicit evaluations and implicit stereotypes are systematically related (Kurdi et al., 2019). However, previous findings remain silent on the processes underlying relations between the two constructs. Drawing on the hypothesis that evaluative and semantic associations mutually constrain each other (Greenwald et al., 2002), the current research investigated (a) whether changes in the valence of semantic attributes associated with a group influence implicit evaluations of that group and (b) whether changes in the valence of a group influence implicit stereotyping of that group.

The Relation Between Stereotypes and Prejudice

Though it is often assumed that group evaluations and stereotypes are related, whether on implicit or explicit measures, this relation varies widely, with some researchers finding only weak relations and others finding stronger relations. For example, in a review of the available evidence at the time, Brigham (1971) found no systematic relation between explicit evaluations and explicit stereotypes about Black people. Esses et al. (1993), however, found the strength of the relation between evaluations and stereotypes to depend on the group studied. In their study, correlations were as high as $r = .61$ when Jewish people were the target group and as low as $r = .37$ when English Canadians were the target group. Similarly, Phills et al. (2018) found that the strength of the relation between evaluations and stereotypes varied for different subgroups, in that evaluations of Black people were related to stereotypes about Black men but not Black women.

Using implicit measures, Amodio and Devine (2006) found no relation between evaluations and stereotyping of Black

[1]University of North Florida, Jacksonville, USA
[2]University of Cologne, Germany
[3]The University of Texas at Austin, USA

*First authorship is shared equally between the first two authors.

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

Adam Hahn, Social Cognition Center Cologne, University of Cologne, Richard-Strauss-Str. 2, 50931 Köln, Germany.
Email: Adam.Hahn@uni-koeln.de

people. In their study, the measure of implicit stereotyping included the neutral attribute labels mental and physical. The authors explained the difference between their and previous findings suggesting that neutral stereotypic content would not be related to evaluations, but valenced stereotypic content would. Consistent with this argument, Rudman and Goodwin (2004) found that implicit evaluations of men and women were related to stereotypes regarding threat and safety. However, they found no relation between implicit evaluations of men and women and stereotypes regarding power and warmth.

Using large samples and archival data, Kurdi et al. (2019) report strong and consistent relations between implicit evaluations and implicit stereotyping. Moreover, in the first study explicitly testing a causal relation between stereotypes and evaluations, Kurdi et al. found that experimentally induced changes in the valence of new fictional groups also changed associations of those groups with positive or negative attributes (stereotypes). The authors explained the inconsistency between their and previous findings stating that relations between evaluations and stereotyping should be stronger on implicit measures than explicit measures, because the latter may be more strongly influenced by cultural knowledge of social stereotypes. However, this interpretation seems difficult to reconcile with the results of other studies, such as the reviewed findings by Amodio and Devine (2006). It is also at odds with arguments in the debate on whether cultural knowledge has a greater impact on implicit as opposed to explicit measures (e.g., Devine, 1989; Gawronski et al., 2008; Payne et al., 2017).

The Balance-Congruity Principle

The reviewed evidence indicates that relations between evaluations and stereotypes vary considerably across studies, target groups, and stereotypical attributes. However, the mental processes underlying relations between the two constructs are still unclear. In an attempt to fill this gap, we draw upon the balance-congruity principle of Greenwald et al.'s (2002) unified theory (UT) of social cognition. The principle is based on a definition of *shared first-order link*, stating that two nodes have a shared first-order link when each of the two nodes is linked to the same third node. For example, a group (e.g., Americans) and an attribute (e.g., strong) share a first-order link if they both are linked to positive evaluations. Expanding on this definition, the balance-congruity principle states that "when two unlinked or weakly linked nodes share a first-order link, the association between the two should strengthen" (Greenwald et al., 2002, p. 6). Thus, an association should form between a group and an attribute when they both share a first-order link to a third node like positive evaluations. Applied to the current question, an important implication of the balance-congruity principle is that group–valence associations, group–attribute associations, and attribute–valence associations should mutually constrain each other to maintain a balanced triad. Specifically, implicit evaluations of a particular group (reflecting group–valence associations according to UT) should be constrained by the valence of stereotypical attributes (attribute–valence associations) that are associated with the group (group–attribute associations). Conversely, implicit stereotypes about a particular group (reflecting group–attribute associations according to UT) should be constrained by the valence of the group (group–valence associations) and the valence of the stereotypical attributes (attribute–valence associations).

Although earlier work guided by UT has focused predominantly on correlational predictions (Greenwald et al., 2002), the causal relations implied by the balance-congruity principle can also be tested experimentally (e.g., Dunham, 2013; Phills et al., 2019). Specifically, the balance-congruity principle suggests that, if a group is stereotypically associated with a specific Attribute A and the valence of Attribute A changes, the target group's valence should also change in the same direction. Conversely, if the evaluation of a group changes, then this group should become more strongly associated with attributes of the same valence and less strongly with attributes of opposing valence. Hence, in contrast to the lack of theorizing on the relation between stereotypes and prejudice presented above, the balance-congruity principle leads to clear predictions about how these constructs should mutually influence each other.

The Current Research

Expanding on the implications of the balance-congruity principle, the current research investigated whether (a) experimentally induced changes in the valence of semantic attributes associated with a group (stereotypes) influence implicit evaluations of that group (Experiments 1 and 2), and (b) experimentally induced changes in the valence of a group (prejudice) influence implicit stereotyping of that group (Experiments 3 and 4). Across the four studies, we tested each prediction with both fictional (Experiments 1 and 3) and real groups (Experiments 2 and 4).

The sample sizes in the first two studies were based on a meta-analysis of evaluative conditioning (EC), which revealed an average effect size of $d = .52$ (Hofmann et al., 2010). Based on this effect size, we aimed to recruit 120 participants in Experiments 1 and 2, which provides a power of 80% in obtaining a significant difference between two independent means at the $p = .05$ level in a two-tailed test (Faul et al., 2007). Because the observed effect sizes in the first two studies were somewhat smaller than expected ($\sim d = .46$), we aimed for a sample size of 150 participants in Experiments 3 and 4, again corresponding to a power of 80%. The data for each study were collected in one shot without intermittent statistical analyses. We report all measures, all conditions, and all data exclusions. All data, materials, and statistical analysis files are publicly available at https://osf.io/kjz98/.

Experiment 1

The purpose of Experiment 1 was to test whether changing the evaluations of attributes previously associated with a group would lead to changes in evaluations of the group itself. Specifically, we used a sensory preconditioning procedure (Walther, 2002) to investigate whether experimentally induced changes in the valence of neutral attributes that have previously been associated with a novel group influences implicit evaluations of that group. Two fictional groups were repeatedly paired with nonword attributes. Afterward, participants' evaluative representation of the pre-associated attributes was manipulated via EC by repeatedly pairing these attributes with either pleasant or unpleasant images (see Hofmann et al., 2010). In this procedure, the groups themselves were never directly paired with pleasant or unpleasant stimuli. Hence, any differences in implicit evaluations of the groups are the result of their previously established associations with the attributes and the newly acquired valence of the attributes (Walther, 2002).

Method

Participants and design. One hundred twenty-one undergraduates at the University of Western Ontario in Canada participated in the study for course credit. Data from four participants were excluded from the analyses. One participant failed to complete all measures; one participant completed the experiment twice; one participant reported ignoring the instructions; and for a fourth participant the images did not load during the experiment. The remaining 117 participants (40.2% male, median age = 20) were randomly assigned to one of two conditions in which fictional groups were associated with neutral attributes that were later paired with pleasant or unpleasant images. Specifically, half of the participants learned to associate Novel Group 1 with a neutral attribute that was later paired with pleasant images and Novel Group 2 with a neutral attribute that was later paired with negative images. The remaining half learned the opposite pairing.

Procedure. To discourage participants from spontaneously drawing inferences about how one task was connected to another, participants were informed that they would complete a series of separate studies that had been combined to make better use of the participant pool (in reality, all tasks were part of the same experiment). In the first "study," participants were told to imagine they were scientists who had discovered an alien species on a distant planet. The purpose of this task was for participants to learn to associate the two groups with novel neutral attributes (i.e., *axpart* vs. *fronded*; Richards & Blanchette, 2004).[1] In the second "study," participants were asked to respond to attribute–image pairs consisting of the attributes from the first "study" and a pleasant or unpleasant image. The task was designed as an EC procedure to create positive associations with one of the initially neutral attributes and negative associations with the other initially neutral attribute. In the third "study," participants completed an implicit association test (IAT; Greenwald et al., 1998) designed to assess implicit evaluations of the two alien groups. Afterward, participants completed a manipulation check for the EC task that measured their evaluations of the previously neutral novel attributes. Finally, participants completed demographic questions before being thanked and debriefed.

Materials

Attribute learning task. Participants were asked to learn the attributes associated with two novel social groups, red and yellow aliens that ostensibly lived on a distant planet. The alien groups were never named to ensure that the attributes were the only semantic information associated with each group, and the attributes were pronounceable nonwords (i.e., *axpart* vs. *fronded*; Richards & Blanchette, 2004). Participants' task was to identify the "attributes" associated with two groups of aliens during a task in which attributes and members of the two groups would be presented together repeatedly on the computer screen. On each trial of the task, a member of one alien group and either the nonword *axpart* or the nonword *fronded* were presented beside one another in the center of the computer screen. The nonwords were each presented to the right or left of the aliens on equal amounts of trials. For half of the participants, members of Novel Group 1 were paired with *axpart* 20 times and members of Novel Group 2 were paired with *fronded* 20 times. For the remaining half of participants, these pairings were reversed. Each pairing appeared for 1,000 ms followed by a blank screen for 1,000 ms before the presentation of the next pair (Gregg et al., 2006).

Evaluative conditioning task. Participants were presented with 40 pairings of attributes from the previous task (i.e., *axpart* vs. *fronded*) along with positive and negative images (i.e., puppies vs. skulls; see Lang et al., 2008). For half of the participants, the nonword *axpart* was always paired with positive images and the nonword *fronded* was always paired with negative images. For the remaining participants, the pairings were reversed. To encourage participants to pay attention to the pairings, all participants were instructed to press the "E" key when the nonword appeared on the left side of the screen and to press the "I" key when the nonword appeared on the right side of the screen. The pairings remained on the screen for a full 1,000 ms regardless of how quickly participants responded to the stimuli. After each pairing, participants were presented with either a blank screen (correct response) or an "X" (incorrect response), each for 1,000 ms, before the presentation of the next pairing. Whenever participants did not respond within 1,000 ms of the onset of the stimuli, they were presented with the message *Please try to respond faster*.

Implicit group evaluations. Implicit evaluations of the two novel groups were assessed using the IAT (Greenwald et al., 1998). Divided into five blocks, the task required participants to categorize members of the two novel groups as well as six pleasant words and six unpleasant words (see https://osf.io/kjz98/) as quickly as possible. In the first and fourth block, participants categorized members of the two novel groups; in the second block, participants categorized the pleasant and unpleasant words. In the third block (i.e., initial combined block), participants categorized members of the two novel groups along with the pleasant and unpleasant words, such that the same response key was used for members of Novel Group 1 and pleasant words and another key for members of Novel Group 2 and unpleasant words. These key mappings were reversed in the fifth block (Novel Group 1 + bad, Novel Group 2 + good; reversed combined block). The order of the two combined blocks was counterbalanced across participants. Each combined block included a total of 72 trials, and an intertrial window of 1,250 ms after correct responses. When participants made an incorrect response, they were presented with a blank screen for 500 ms, an "X" in the center of the screen for 250 ms, then another blank screen for 500 ms before the presentation of the next trial. Participants did not receive an opportunity to correct incorrect responses. We used the *D*-600 algorithm to calculate IAT scores of implicit evaluations (Greenwald et al., 2003).

Manipulation check. To confirm the effectiveness of the EC manipulation, participants rated how pleasant or unpleasant they found the nonwords *axpart* and *fronded* on 7-point rating scales ranging from 1 (*very unpleasant*) to 7 (*very pleasant*). Ratings of the two attributes *axpart* and *fronded* were combined in a single index by calculating a difference score reflecting greater relative liking of *axpart* over *fronded*.

Results

Manipulation check. Consistent with the intended effect of the EC manipulation, participants who were shown pairings of *axpart* with positive images and *fronded* with negative images showed a stronger preference for *axpart* over *fronded* (M_{diff} = 0.56, *SD* = 2.06) than participants who were shown reversed pairings (M_{diff} = −0.24, *SD* = 1.57), *F*(1, 115) = 5.57, *p* = .020, η_p^2 = .046, 90% confidence interval [CI] [.004, .121].

Implicit group evaluations. IAT scores were calculated such that higher scores reflect a greater implicit preference for Novel Group 1 over Novel Group 2. To test the impact of the experimentally induced changes in the valence of group attributes on implicit group evaluations, IAT scores were submitted to a 2 (Group Attribute Valence: Novel Group 1 associated with positive attribute + Novel Group 2 associated with negative attribute vs. Novel Group 2 associated with positive attribute + Novel Group 1 associated with negative attribute) × 2 (IAT Block Order: Novel Group 1 + positive first vs. Novel Group 2 + positive first) ANOVA. Means for all conditions are presented in Table 1. The ANOVA revealed a theoretically uninteresting main effect of IAT Block Order, *F*(1, 113) = 96.02, *p* < .001, η_p^2 = .46, 90% CI [.347, .545], and, more importantly, a significant main effect of Group Attribute Valence, *F*(1, 113) = 6.50, *p* = .012, η_p^2 = .054, 90% CI [.007, .134]. Consistent with predictions, participants showed a greater implicit preference for Novel Group 1 over Novel Group 2 when the attribute previously associated with Novel Group 1 was later associated with a positive image and the attribute previously associated with Novel Group 2 was later associated with a negative image (*M* = .01, *SD* = .43), compared with the reverse pairing (*M* = −.14, *SD* = .37). There was no significant interaction between Group Attribute Valence and IAT Block Order, *F*(1, 113) = 0.29, *p* = .589, η_p^2 = .003, 90% CI [.000, .039].

Discussion

Using a procedure adopted from research on sensory preconditioning (Walther, 2002), Experiment 1 found that changes in the valence of formerly neutral attributes produced corresponding changes in implicit evaluations of novel groups that had been pre-associated with these attributes. These results support the prediction that changing the valence of an attribute associated with a group changes implicit evaluations of that group. However, one limitation of this study is that, although we told participants that the nonword attributes are descriptive of the alien species, they were not part of a rich semantic network with links to evaluative associations, as is often the case for real stereotypes. Experiment 2 tested if we could obtain similar effects with real groups and attributes.

Experiment 2

The purpose of Experiment 2 was to replicate the findings of Experiment 1 using real groups and attributes. Toward this end, participants were instructed to write about the reasons and benefits for regular people to try maintaining a high level of either (a) physical fitness or (b) mental fitness. Thinking of the benefits of physical versus mental fitness was assumed to temporarily bolster positive evaluations of the attributes *athletic* or *intelligent*, respectively. Afterward, participants completed an IAT designed to measure implicit evaluations of athletes (a group stereotypically associated with the attribute *athletic*) and scientists (a group stereotypically associated with the attribute *intelligent*). Expanding on the findings in Experiment 1, we expected that participants who were instructed to think about the benefits of being athletic should show a greater preference for athletes over scientists compared with participants who were instructed to think about the benefits of being intelligent.

Table 1. IAT Scores by Experiment, Condition, and IAT Block Presentation Order.

Experimental Manipulation	Condition	IAT block presentation order 1		IAT block presentation order 2	
		M	*SD*	*M*	*SD*
Experiment 1, DV: Evaluation of Novel Group 1 over 2					
EC pairing	Group 1 with later-positive attribute + Group 2 with later-negative attribute	.29	.33	–.28	.32
	Group 2 with later-positive attribute + Group 1 with later-negative attribute	.12	.29	–.39	.25
Experiment 2, DV: Evaluation of athletes over scientists					
Writing Task	Benefits of physical fitness	.29	.42	*n/a*	
	Benefits of mental fitness	.06	.60	*n/a*	
Experiment 3, DV: Stereotyping of Novel Group 1 over 2 as intelligent over aggressive					
EC pairing	Novel Group 1 with positive	–.03	.35	–.19	.35
	Novel Group 2 with positive	–.22	.39	–.27	.42
Experiment 4, DV: Stereotyping of African Americans as athletic over aggressive					
EC pairing	African American with positive	.21	.37	.10	.48
	African American with negative	.01	.36	–.07	.27

Note. In Experiment 1, IAT block presentation order 1 refers to Novel Group 1 + positive first and IAT block presentation order 2 refers to Novel Group 2 + positive first. In Experiment 3, IAT block presentation order 1 refers to Novel Group 1 + intelligent first and IAT block presentation order 2 refers to Novel Group 2 + intelligent first. In Experiment 4, IAT block presentation order 1 refers to African American + aggressive first and IAT block presentation order 2 refers to African American + athletic first. DV = dependent variable; IAT = implicit association test; EC = evaluative conditioning.

Method

Participants and design. One-hundred twenty-nine (19.4% male, median age = 21) participants were recruited on campus at the University of Cologne in Germany (96.9% university students) via flyers, email lists, and direct approach to participate in a 10 to 15-min study on "leisure time activities" in exchange for a chocolate bar and optional experimental credit when applicable. The study consisted of a two-condition (physical fitness vs. mental fitness) between-subjects design with implicit evaluations of athletes versus scientists as the dependent variable.

Attribute evaluation manipulation. After signing informed consent, participants read an introductory page explaining that we were piloting materials for use in later studies. In keeping with the cover story, the first page began with the questions *How do people spend their time? What motivates people to make use of their time in a certain way?* To manipulate the valence of the attributes *athletic* and *intelligent*, the line below either read *physical activity* or *mental exercise and general knowledge* followed by two paragraphs that included further information about the upcoming task (sentences that differ between conditions are highlighted with mental fitness condition in brackets):

> In this study we are interested in why people decide to dedicate their time to physical activity [mental exercises and to improving their knowledge and education]. Some people dedicate some or a lot of time, others dedicate little or no time to, e.g., training their muscles and physical endurance, or to becoming really good at a specific athletic discipline [crossword puzzles, Sudokus, playing strategy games, or to reading news and non-fiction books]. We are collecting ideas about why people dedicate their time to physical activity [these kinds of mental exercises]. What do you think? Many people have only limited time for various activities. Why is it important to people to be physically active [educated and well-read, or good at mental exercises]? What reasons might there be? Please write down three reasons why you personally think that people want to be athletic and physically fit [mentally fit and educated], and dedicate their time accordingly. We will ask you to explain your answers afterwards.

After participants identified three reasons, the next page asked them to elaborate on those reasons by writing down their thoughts in a large essay box. The rationale of this exercise was to make participants generate positive consequences, and thus activate positive associations with being athletic and physically fit (attributes stereotypically associated with athletes) or with being intelligent and educated (attributes stereotypically associated with scientists).

Implicit group evaluations. To test the effect of the writing task on implicit evaluations of groups associated with the attributes from the writing task, participants completed an IAT designed to measure implicit evaluations of athletes and scientists (Greenwald et al., 1998). The stimuli used to represent each of the two groups were 10 pictures of young adult males who engaged in activities and wearing attire that is typical for the two categories. The 10 athlete pictures showed men long-jumping, weightlifting, running, playing basketball, biking, hurdling, rock climbing, playing soccer or handball, and swimming. The 10 scientist pictures showed men lecturing in front of blackboards with formulas (four different pictures), in front of computers (2), in lab coats (3), or holding cables (1). Although we did not formally pretest the images, the pictures were chosen to show men of approximately similar ages. Participants were asked to sort these pictures as either representing a scientist or an athlete. On the two combined blocks of the IAT, participants completed 60 trials sorting the pictures simultaneously with 10 positive and 10 negative words. In the initial combined block, responses to scientists were mapped onto the same key as positive words and responses to athletes were mapped onto the same key as negative words. In the reversed combined block, responses to scientists were mapped onto the same key as negative words and responses to athletes were mapped onto the same key as positive words.[2] IAT scores were calculated using the *D*-600 algorithm utilized in Experiment 1. Higher scores indicate more positive evaluations of athletes compared to scientists.

Explicit group evaluations. For exploratory purposes, the current study also included a measure of explicit group evaluations, presented after the IAT. Participants indicated their explicit preference for scientist as opposed to athletes on a 7-point rating scale ranging from 1 (*considerably more positive toward athletes*) to 7 (*considerably more positive toward scientists*). Responses on the measure of explicit evaluations were reverse-scored, such that higher scores indicate a greater preference for athletes over scientists (as in the IAT).

Results

Manipulation check. To check whether the manipulations had indeed made participants generate positive thoughts about the attributes in general rather than thoughts about specific exemplars excelling at these attributes (i.e., athletes or scientists), two research assistants coded all individual answers in two different random orders (interrater reliability, Kendall's Tau = .84). Both coders agreed that only two out of the 502 responses[3] mentioned a specific exemplar (Alfred Hitchcock and a participant's grandfather, both in the intelligence condition). All other responses described the benefits of a physically active or intellectually challenging lifestyle for nonspecific regular people, as intended.[4]

Implicit group evaluations. A one-way ANOVA comparing the IAT scores in the two conditions replicated the main finding of Experiment 1 (see Table 1). Participants who wrote about the benefits of being physically fit showed a greater implicit preference for athletes over scientists (M = .29, SD = .42) than participants who wrote about the benefits of being mentally fit (M = .06, SD = .60), $F(1, 127) = 5.99$, $p = .016$, $\eta_p^2 = .045$, 90% CI [.005, .116].

Explicit group evaluations. Controlling for condition, explicit and implicit preference for athletes over scientists showed a significant positive partial correlation ($r = .42$, $p < .001$). An exploratory one-way ANOVA on explicit preference scores revealed a trend in the same direction as the IAT results ($M_{\text{physical}} = 4.85$, $SD = 1.14$; $M_{\text{mental}} = 4.47$, $SD = 1.41$), but the difference was not statistically significant, $F(1, 127) = 2.80$, $p = .097$, $\eta_p^2 = .022$, 90% CI [.000, .079].

Discussion

Experiment 2 replicated the main finding of Experiment 1 using real instead of fictitious groups. Participants instructed to write about the benefits of physical fitness for regular people showed a greater implicit preference for athletes over scientists than participants instructed to write about the benefits of mental fitness. Together, the two experiments suggest that changes in the valence of a given attribute lead to corresponding changes in implicit evaluations of groups associated with that attribute. For exploratory purposes, the current study also included a measure of explicit group evaluations. Although this measure showed a trend in the same direction as the measure of implicit group evaluations, the difference between conditions was not statistically significant. We will return to this finding in the General Discussion.

Experiment 3

To investigate the reverse direction of the bidirectional relation between evaluation and stereotyping, Experiment 3 tested whether changes in the valence of a group lead to changes in implicit stereotyping. Toward this end, participants in Experiment 3 learned to associate each of two novel groups with both a positive and a negative trait at the same time (i.e., *intelligent* and *aggressive*). Hence, both groups were associated equally with both traits. We chose to use valenced attributes rather than neutral nonwords in this experiment to reduce the number of tasks participants completed and to increase the possibility that the first task would in fact lead to semantic storage of the associations between the new groups and the attributes. Next, the valence of the two groups was manipulated via EC by repeatedly pairing one of them with positive images and the other one with negative images (Hofmann et al., 2010). Finally, participants completed an IAT designed to measure implicit stereotyping along the two trait dimensions of the attribute learning task (i.e., *intelligent* and *aggressive*). Drawing on the balance-congruity principle, we expected stronger implicit stereotyping of a given group as being intelligent (vs. aggressive) when the group was paired with positive images than when it was paired with negative images. Conversely, implicit stereotyping of a given group as being aggressive (vs. intelligent) should be stronger when the group was paired with negative images than when it was paired with positive images.

Experiment 3 bares some similarity with a recent study by Kurdi et al. (2019), who showed that an EC procedure associating a novel group with positive (vs. negative) valence led to associations of this group with *American* (i.e., a positive attribute for their American participants) as opposed to *foreign* (i.e., a negative attribute for their American participants). Yet, different from the procedure in the current study, Kurdi et al. did not pre-associate the novel groups with positive and negative attributes. Because neither group had previously been associated with the attribute "American," their findings may speak more to how evaluative representations of social groups constrain the acquisition of stereotypic knowledge about a group than to how newly formed evaluative representations change existing stereotypic representations. The present study first had participants associate two attributes with two alien groups to investigate whether changes in the valence of a group would change existing associations. In addition, Kurdi et al. (2019) did not offer a specific theoretical explanation for when and why stereotypes and prejudice should be related. Experiments 3 and 4 aim to fill this gap by presenting a specific theoretical model based on the balance-congruity principle about when and why changes in evaluations can change existing stereotypes, in addition to the reverse causal patterns demonstrated in Experiments 1 and 2. Experiment 4 will further complement these findings with an investigation with real groups. Otherwise the design and goals of Experiment 3 were consistent with Kurdi et al.'s (2019) Study 2.

Method

Participants and design. One hundred fifty (39.3% male, median age = 36) participants (17 undergraduates at the University of North Florida and 133 Mechanical Turk workers) completed the experiment online at a location of their choice in exchange for experimental credit (undergraduates) or $1 (MTurk).[5] All participants were randomly assigned to one of two EC conditions (Novel Group 1 with positive images and Novel Group 2 with negative images vs. Novel Group 1 with negative images and Novel Group 2 with positive images) in a between-subjects design.

Procedure. Upon logging into the experiment website, participants were informed that they would be participating in a series of separate and unrelated tasks. As in Experiment 1, the first task instructed participants to imagine they were scientists who had discovered an alien species on a distant planet and attempt to identify the "attributes" associated with each alien group during a task in which attributes and members of the two groups would be presented together on the computer screen. Different from Experiment 1, the purpose of this task was for participants to simultaneously associate each group with one positive attribute (i.e., *intelligent*) and one negative attribute (i.e., *aggressive*). As in Experiment 1, the second task required participants to respond to pairs of stimuli presented in the center of the screen. Unlike Experiment 1, however, these pairs of stimuli consisted of the alien group members presented during the first task along with positive and negative images. The purpose of this task was to create positive associations with one of the two groups and negative associations with the other. In the third task, participants completed an IAT designed to assess implicit stereotyping of the two groups along the two attributes from the first task (i.e., *intelligent* and *aggressive*). As in Experiment 2, participants also answered exploratory explicit questions about their stereotyping of the novel groups after completing the IAT, before being thanked and debriefed.

Materials

Attribute learning task. As in Experiment 1, participants were asked to learn attributes associated with two novel groups. Toward this end, traits related to both a positive and a negative attribute were presented an equal number of times with each alien group, such that both groups were associated equally with the two attributes. The positive attribute was *intelligent* and the presented trait words were *intelligent, brainy, educated, smart, genius*, and *clever*. The negative attribute was *aggressive* and the presented trait words were *aggressive, hostile, angry, combative, threatening*, and *violent*. Forty group–attribute pairings were presented to participants in random order, with each alien group being presented 10 times with traits related to each attribute. Each image–attribute pair was presented onscreen for 1,000 ms followed by a blank screen for 1,000 ms.

Evaluative conditioning task. Participants were presented with 40 pairings of the aliens along with positive and negative images (i.e., puppies vs. skulls). For half of the participants, the positive image was paired with Alien Group 1 and the negative image was paired with Alien Group 2, while pairing was reversed for the other group. All other procedural details were similar to the EC procedure employed in Experiment 1.

Implicit stereotyping. Implicit stereotyping of the two novel (alien) groups was assessed using an IAT (Greenwald et al., 1998). Participants were asked to categorize members of the two groups as well as five traits related to *intelligence* (the same traits used in the attribute learning task, except *clever*) and five traits related to *aggression* (the same traits used in the attribute learning task, except *violent*) as quickly as possible. In one of the two combined blocks (60 trials each), the stimuli were paired such that Novel Group 1 shared a response key with *intelligence* words and Novel Group 2 shared a response key with *aggression* words. The other combined block used the reversed mapping. The order of the two combined blocks was counterbalanced across participants. All other procedural details of the IAT were identical to Experiment 1. IAT scores were again calculated using the *D*-600 algorithm (Greenwald et al., 2003). Higher scores represent stronger associations with intelligence and weaker associations with aggression for Novel Group 1 as opposed to Novel Group 2.

Explicit stereotyping. A measure of explicit stereotyping was again included for exploratory purposes after the IAT. Participants rated the extent to which they associated the six attributes related to aggression and the six attributes related to intelligence used on the IAT with each novel group on 7-point rating scales ranging from 1 (*strongly associated with red aliens*) to 7 (*strongly associated with yellow aliens*). Participants' trait ratings were averaged into one score for intelligence (Cronbach's $\alpha = .96$) and one score for aggression (Cronbach's $\alpha = .94$). We then subtracted aggression scores from intelligence scores to create a single measure of explicit stereotyping such that higher scores correspond directionally to the measure of implicit stereotyping.

Results

Implicit stereotyping. To investigate changes in implicit stereotyping of the two groups as a result of changes in the valence of those groups, we conducted a 2 (Evaluative Conditioning: Novel Group 1 Positive vs. Novel Group 2 Positive) × 2 (IAT Block Order: Novel Group 1 + intelligent first vs. Novel Group 2 + intelligent first) ANOVA (see Table 1 for all means). The analysis revealed a significant main effect of Evaluative Conditioning, $F(1, 146) = 4.87$, $p = .029$, $\eta_p^2 = .032$, 90% CI [.002, .092], indicating that participants who were shown pairings of Novel Group 1 with positive images and Novel Group 2 with negative images tended to show higher scores on the implicit stereotyping index ($M = -.11$, $SD = .36$) compared with participants who were shown pairings of Novel Group 1 with negative images and Novel Group 2 with positive images ($M = -.25$, $SD = .40$). In other words, seeing pairings of Novel Group 1 with positive images and Novel Group 2 with negative images led to stronger associations of Novel Group 1 with intelligence over aggression, but to stronger associations of Novel Group 2 with aggression over intelligence, whereas the reverse pattern could be observed when participants had seen the opposite pairing. Neither the main effect of IAT Block Order, $F(1, 146) = 3.057$, $p = .083$, $\eta_p^2 = .021$, 90% CI [.000, .072], nor the interaction between Evaluative Conditioning and IAT Block Order were statistically significant, $F(1, 146) = 0.67$, $p = .415$, $\eta_p^2 = .005$, 90% CI [.000, .019].

Explicit stereotyping. Controlling for experimental condition, explicit and implicit stereotyping scores showed a significant positive partial correlation ($r = .26$, $p = .002$). Exploratory analyses revealed a trend in the same direction as the IAT results ($M = -0.93$, $SD = 2.26$ vs. $M = -1.48$, $SD = 2.25$), but the difference between conditions did not reach statistical significance, $F(1, 148) = 2.22$, $p = .139$, $\eta_p^2 = .015$, 90% CI [.000, .062].

Discussion

Experiment 3 provides evidence that changes in the valence of a group lead to corresponding changes in implicit stereotyping. Although both target groups were associated equally with the same positive and negative traits (i.e., intelligent and aggressive), each group became more strongly associated with the trait that matched the subsequently conditioned valence of the group. Compared with participants for whom Novel Group 1 was paired with a negative image and Novel Group 2 was paired with a positive image, participants with the reverse group–image pairings associated Novel Group 1 more strongly with the positive trait than the negative trait and Novel Group 2 more strongly with the negative than the positive trait. As in Experiment 2, an exploratory explicit measure showed a similar mean pattern, but the difference between conditions was not statistically significant.

Experiment 4

Experiment 4 aimed to replicate the findings of Experiment 3 using a real social group: African Americans. Toward this end, participants' evaluative representation of African Americans was manipulated by means of an EC task that paired African American faces with positive images and European American faces with negative images, or vice versa (see Olson & Fazio, 2006). Expanding on the results of Experiment 3, we tested whether the degree of implicit stereotyping of African Americans as either athletic (positive attribute) or aggressive (negative attribute) depended on whether African Americans had been paired with positive or negative images in the EC task.

Methods

Participants and design. One hundred fifty three undergraduates at the University of North Florida completed the experiment online in exchange for course credit. Data from 19 participants were excluded from analyses because they were African American, and from four participants because they did not disclose their race. The remaining 130 participants (106 White, 15 Hispanic, 7 Asian, 1 Native American, 1 Indian, 13.8% male, median age = 19) were randomly assigned to conditions in which (a) African American faces were paired with positive images and European American faces were paired with negative images or (a) African American faces were paired with negative images and European American faces were paired with positive images.

Procedure. Upon logging in to the experiment website, participants completed an EC procedure designed to associate African Americans with positive images and European Americans with negative images, or vice versa. Next, participants completed a Single-Category IAT (SC-IAT; Karpinski & Steinman, 2006) designed to assess the strength of association between African Americans and the attributes *aggressive* and *athletic*, before answering demographics questions and being presented with a debriefing screen.

Materials

Evaluative conditioning task. Participants completed an EC task in which positive and negative images were paired with two social categories. To increase participants' motivation to learn the pairings, participants were told that they would be asked to complete a memory test later in the study (in reality, there was no such memory test). On each trial of the task, participants were shown a head-and-shoulders photograph of a person for 1,000 ms followed by a positive or negative image for 1,000 ms and then a prompt to press any key to view the next pairing. Five photos of African Americans, five photos of European Americans, five positive images, and five negative images were shown to participants during the task. In the African American–positive/European American–negative condition, photos of African Americans were always followed by positive images and photos of European Americans were always followed by negative images. The group–valence pairings were reversed in the African American–negative/European American–positive condition. In total, participants viewed 80 pairings and were given breaks after completing 25%, 50%, and 75% of the task.

Implicit stereotyping. After completion of the EC task, participants were informed that they would complete a number of unrelated tasks before the memory test. The first of these was a SC-IAT (Karpinski & Steinman, 2006) designed to assess the implicit stereotyping of African Americans in terms of the attributes *aggressive* and *athletic*. In the first block of this task, participants practiced categorizing trait words presented in the center of the screen. Participants were asked to press the "E" key when an aggressive trait word (*aggressive, threatening, violent, hostile, fierce, offensive*) was presented on the screen, and the "I" key when an athletic trait word (*athletic, active, energetic, fit, sporty, agile*) was presented. Participants then completed two combined blocks in which they responded to photos of African Americans not used in the previous tasks (3 women, 3 men) in addition to the athletic and aggressive trait words. In one of the two combined blocks, participants were instructed to respond to photos of African Americans and aggressive traits using the "E" key and to respond to athletic traits using the "I" key. In the other combined block, this pairing was reversed (African American + athletic vs. aggressive). Each combined block consisted of 40 trials. The order of the two combined blocks was counterbalanced across participants. When participants made a correct response, they were presented with a blank screen for 1,000 ms before the presentation of the next trial. When participants made an incorrect response, they were presented with a blank screen for 100 ms followed by the presentation of a red "X" in the center of the screen for 800 ms, and then the presentation of a blank screen for 100 ms before the start of the next trial. IAT scores were calculated in line with the procedures in Experiments 1 to 3. Higher scores reflect stronger implicit stereotyping of African Americans as athletic compared with aggressive.

Explicit stereotyping. For exploratory purposes, participants were asked to rate how strongly they associated traits related to *athletic* and *aggressive* with African Americans on a 7-point scales ranging from 1 (*not at all*) to 7 (*very strongly*). The adjectives for the two trait dimensions were the same that were used in the IAT. We then subtracted the mean score for the six *aggressive* items (Cronbach's α = .94) from the mean score for the *athletic* items (Cronbach's α = .95), such that higher scores indicate stronger explicit stereotyping of African Americans as athletic compared with aggressive.

Results

Implicit stereotyping. Means for all conditions are presented in Table 1. A 2 (Evaluative Conditioning: African American–positive vs. African American–negative) × 2 (IAT Block Order: African American + aggressive first vs. African American + athletic first) ANOVA revealed a significant main effect of Evaluative Conditioning, such that participants who were shown pairings of African Americans with positive images and European Americans with negative images associated African Americans more strongly with *athletic* versus *aggressive* (M = .15, SD = .43) compared with participants who were shown pairings of African Americans with negative images and European Americans with positive images (M = −.027, SD = .32), $F(1, 126) = 7.53$, $p = .007$, $\eta_p^2 = .056$, 90% CI [.009, .132]. Neither the main effect of IAT Block Order, $F(1, 126) = 1.92$, $p = .169$, $\eta_p^2 = .015$, 90% CI [.000, .067], nor the interaction between Evaluative Conditioning and IAT Block Order were statistically significant, $F(1, 126) = 0.02$, $p = .886$, $\eta_p^2 < .001$, 90% CI [.000, .013].

Explicit stereotyping. Controlling for experimental condition, implicit stereotyping of African Americans did not show a significant partial correlation with explicit stereotyping, $r = .12$, $p = .161$. Exploratory analyses revealed a trend in the same direction as the IAT results ($M = -1.30$, $SD = 1.70$ vs. $M = -1.66$, $SD = 1.45$), but this difference was not statistically significant, $F(1, 128) = 1.66$, $p = .200$, $\eta_p^2 = .013$, 90% CI [.000, .062].

Discussion

Expanding on the main finding of Experiment 3, Experiment 4 provides evidence that changes in the valence of a real social group influences implicit stereotyping of that group. In the current study, African Americans were more strongly associated with the trait *athletic* and less strongly with the trait *aggressive* after repeated pairings of African Americans with positive images. In contrast, African Americans were more strongly associated with the trait *aggressive* and less strongly with the trait *athletic* after repeated pairings of African Americans with negative images. A similar pattern emerged on explicit measures, but the difference between conditions was not statistically significant.

General Discussion

Research on the relation between prejudice and stereotyping has revealed mixed findings, with some authors finding no relation at all (e.g., Amodio & Devine, 2006) and other authors reporting large correlations and evidence for causal relations (e.g., Kurdi et al., 2019). Yet, other researchers found relations that vary as a function of target group (e.g., Esses et al., 1993; Phills et al., 2018) and stereotypical attributes (Rudman & Goodwin, 2004). In the current work, we drew upon the balance-congruity principle (Greenwald et al., 2002) to generate testable hypotheses about bidirectional causal relations between implicit evaluations and implicit stereotyping of social groups. Specifically, we predicted that changes in the valence of an attribute associated with a group should change implicit evaluations of a group. Vice versa, changes in the valence of a group should change the degree to which it is associated with positive or negative attributes. These predictions were confirmed in four studies for novel and real social groups.

In Experiments 1 and 2, changing the valence of attributes changed implicit evaluations of groups associated with those attributes. In Experiment 1, we demonstrated this by pre-associating novel groups (aliens) with one of two novel attributes (*axpart* and *fronded*). Participants then learned to associate one of these attributes with positive valence and the other one with negative valence. Confirming predictions, participants showed implicit evaluations of the novel alien groups in line with the newly learned valence of the attributes those groups were pre-associated with. In Experiment 2, we demonstrated the same effect on real groups. Participants were asked to think about the benefits of leading a mentally or physically active lifestyle in one's leisure time as a manipulation aimed at bolstering positive evaluations of the attributes *athletic* versus *intelligent*. Again, in line with predictions, they later showed more favorable implicit evaluations of athletes as opposed to scientists—groups stereotyped as athletic versus intelligent—as a function of whether they had written about the benefits of being mentally or physically fit.

In Experiments 3 and 4, we demonstrated the opposite causal path. Here, we showed that changing the valence of social groups leads to stronger stereotyping of those groups on attributes that match the new valence of the groups. In Experiment 3, we demonstrated that novel alien groups who were conditioned to be associated with positive or negative valence were then stereotyped more strongly as intelligent or aggressive–positive and negative attributes they had been pre-associated with to equal degrees. Turning to real groups, Experiment 4 showed that an EC procedure conditioning participants to associate more positive or negative valence with African Americans later showed greater stereotyping of African Americans as athletic (positive) or aggressive (negative)–attributes that are known to be associated with African Americans in general. Together, these findings provide support for a bidirectional causal relation between implicit prejudice and implicit stereotypes.

Theoretical Implications

The predictions tested here are consistent with Greenwald et al.'s (2002) UT, which suggests that implicit evaluations and implicit stereotypes are rooted in causally related representations. According to the theory's balance-congruity principle, nodes in a semantic network maintain balance in that evaluative associations of two individual nodes remain consistent when those nodes are also associated with each other. Although earlier work guided by UT has focused predominantly on correlational predictions (Greenwald et al., 2002), the causal relations implied by the balance-congruity principles can also be tested experimentally (Dunham, 2013; Phills et al., 2019). Specifically, any change in the association between a semantic attribute and a particular valence should influence group-valence associations to the extent that there is a shared first-order link between the group and the semantic attribute. Conversely, any change in the association between a social group and a particular valence should influence group–attribute associations to the extent that there is a shared first-order link between the attribute and the same valence. As a result of these processes, experimentally induced changes in the valence of semantic attributes associated with a group should have corresponding effects on implicit evaluations of that group, as shown in Experiments 1 and 2. Moreover, experimentally induced changes in the valence of a group should have corresponding effects on implicit stereotyping of that group, as shown in Experiments 3 and 4. Our findings can be interpreted as evidence supporting Greenwald et al.'s UT and its balance-congruity principle.

Another prominent theory of stereotypes and prejudice is Amodio and colleagues' memory systems model (MSM, Amodio & Ratner, 2011). According to the MSM, implicit evaluations and implicit stereotyping are rooted in independent mental representations with distinct neural substrates. With the assumption of independent representations, it seems unclear how the MSM would have predicted the present findings. Although the MSM may be able to accommodate relations between prejudice and stereotyping, it does not suggest clear predictions concerning when and why they should influence each other. As such, our findings require additions to the MSM that specify when to expect causal influences from one construct on the other.

Although we derived our hypotheses from a theory of associative representation that has been designed to capture relations between implicit measures of different contents

(Greenwald et al., 2002), it is worth noting that our findings can also be explained by nonassociative theories assuming that implicit and explicit measures reflect the same underlying propositional representations (e.g., De Houwer, 2014). For example, participants in Experiment 1 may have learned the two propositions *This type of alien is axpart* and *axpart is good/bad*. Responses on the evaluative IAT could simply reflect the inference that, based on the two propositions, *This type of alien is good/bad*. Similar propositions and inferences may underlie the findings of the other three experiments.

It is also worth noting that we did not find strong dissociations between implicit and explicit measures in Experiments 2 to 4 where both types of measurement outcomes were assessed. Although we did not find any significant effects on explicit measures, they always showed patterns of results in the same direction as our implicit measures. In fact, analyses with standardized scores of implicit and explicit measures, and an interaction between condition and type of measure (implicit vs. explicit) included in the model demonstrated only the already-reported main effects of condition, all Fs $>$ 5.39, all ps $<$ .03, and no significant interactions between condition and type of measure, all Fs $<$ 3.30, all ps $>$ .05. Because the explicit measures in Experiments 2 to 4 were always administered after the implicit measures, the null effects on explicit measures could simply reflect a lack of statistical power in showing corresponding results on the later-assessed explicit measures.

Drawing on dual-process models, the observed differences could also be interpreted as evidence that people will not always base their explicit reports on the cognitions reflected in implicit measures (e.g., Fazio & Olson, 2014; Gawronski & Bodenhausen, 2006, 2011). According to these models, explicit evaluations typically show patterns in the same direction as implicit evaluations, unless participants reject the cognitions underlying their implicit evaluations as a basis for explicit judgments (Gawronski & LeBel, 2008; Olson & Fazio, 2006). Hence, some participants may have based their explicit reports on those cognitions, while others may have rejected them. As a result, explicit measures would have shown weaker, but similar effects on average. The current data cannot distinguish between these competing interpretations. More importantly, however, the present findings show a bidirectional causal relation between stereotyping and prejudice on implicit measures as the UT would predict. Future research is needed to disambiguate whether these effects will show dissociations between implicit and explicit measures, and whether the links between memory contents on which they are based are best conceptualized as associative or propositional.

Implications for Previous Research

In addition to explaining when stereotypes and prejudice should causally influence each other, our findings may have implications for why the relations between evaluations and stereotyping have varied in previous findings. Specifically, our findings suggest that the size of the relation between evaluations and stereotypes may depend on the degree to which the valence of the attribute captured in the stereotype matches the valence of the group. For example, the degree to which the stereotype "African Americans are more physical than mental" is related to implicit evaluations of African Americans should vary as a function of how positively or negatively a person evaluates the attribute *physical* as opposed to the attribute *mental*. From this perspective, Amodio and Devine's (2006) findings that evaluations of Black versus White Americans were unrelated to stereotyping of those groups as physical versus mental can be explained by the lack of consensus in the evaluations of the attributes physical or mental. For some participants, the attribute *physical* may invoke concepts of *physical aggression* (negative), whereas for others it may invoke concepts of *athleticism* (positive). Similarly, Esses et al.'s (1993) findings that stereotypes and prejudice are more strongly related for some groups than others may be explained by the fact that the evaluations of the relevant attributes in their study were more or less consistent with evaluations of the different target groups. Finally, Kurdi et al.'s (2019) finding that implicit evaluations and stereotypes are strongly related can be explained by the fact that the valences of the relevant attributes in their studies were consistent with the valences associated with the relevant target groups. Future research is needed to determine whether these hypotheses predict variations in the relation between implicit evaluations stereotyping.

Conclusion

Previous research revealed inconsistent evidence concerning the degree to which stereotypes and prejudice are related. Based on Greenwald et al.'s UT (2002), we hypothesized that the representations underlying the two kinds of biases should mutually influence each other. In line with this assumption, we found that (a) changes in the valence of semantic attributes associated with a group (stereotypes) influenced implicit evaluations (prejudice) toward that group and (b) changes in the valence of a group influenced implicit stereotyping of that group. Hence, although it seems reasonable to treat prejudice and stereotyping as conceptually distinct constructs, our findings suggest that they are causally related.

Acknowledgments

We would like to thank the following research assistants for their help. Experiment 1 and 3: Jasmine Desjardins, Nicole Dryburgh, Elizabeth Harris, Chris Pulford, Serena Ransom, and Ting Zhao for their help with the data collection. Experiment 2: Nicole Hartwich and Timo Luoma for their help with programming; Nicole Hartwich, Timo Louma, Caterina Quante, and Julia ter Jung for data collection; and finally, Isabel Skuplik and Anne Weitzel for coding the responses.

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) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Bertram Gawronski was awarded funding from the Canada Research Chairs program, Grant # 215983.

ORCID iD

Adam Hahn https://orcid.org/0000-0002-2232-4976

Notes

1. An anonymous reviewer alerted us to the fact that *fronded* can be used as the adjective of *frond* ("finely divided leaf," https://www.dictionary.com/browse/fronded). However, because none of us was aware of this, the attributes were successfully used by Richards and Blanchette (2004), and the apparent meaning of *fronded* would not make sense as a descriptive attribute of a social group, we believe that our procedure had the intended effect. These results should nevertheless be interpreted bearing this limitation in mind.
2. Due to a miscommunication between the three authors, the combined blocks of the IAT were not counterbalanced in this study. The scientist–positive and athlete–negative pairings always came first.
3. All 129 participants had four text boxes on four separate screens to write down three reasons and then elaborate, but not all participants used all four text boxes, resulting in 502 individual answers.
4. Eight participants were coded by at least one of the coders as referring to themselves and how their regular lives would improve. Excluding those participants did not change results.
5. The combination of undergraduate and MTurk participants is due to the fact that we initially opened the study to undergraduates at the University of North Florida, but were unable to complete the data collection before the end of the academic term. To expedite the completion of the data collection, we then posted the study on Amazon's Mechanical Turk until we reached the desired sample size of 150 participants. The data were analyzed after we reached the desired sample size without intermittent statistical analyses.

Supplemental Material

Supplemental material is available online with this article.

References

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*, 652–661.

Amodio, D. M., & Ratner, K. G. (2011). A memory systems model of implicit social cognition. *Current Directions in Psychological Science*, *20*, 143–148.

Brigham, J. C. (1971). Ethnic stereotypes. *Psychological Bulletin*, *76*, 15–38.

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, *8*, 342–353.

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

Dunham, Y. (2013). Balanced identity in the minimal groups paradigm. *PLOS ONE*, *8*(12), Article e84205.

Esses, V. M., Haddock, G., & Zanna, M. P. (1993). Values, stereotypes, and emotions as determinants of intergroup attitudes. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (pp. 137–166). Academic Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Fazio, R. H., & Olson, M. A. (2014). The mode model. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 155–171). Guilford Press.

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

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology*, *44*, 59–127.

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology*, *44*, 1355–1361.

Gawronski, B., Peters, K. R., & LeBel, E. P. (2008). What makes mental associations personal or extra-personal? Conceptual issues in the methodological debate about implicit attitude measures. *Social and Personality Psychology Compass*, *2*, 1002–1023.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*, 3–25.

Greenwald, A. G., McGhee, D., & Schwartz, J. (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*, 481–481.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*, 1–20.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*, 390–421.

Karpinski, A., & Steinman, R. B. (2006). The Single Category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*, 16–32.

Kurdi, B., Mann, T. C., Charlesworth, T. E., & Banaji, M. R. (2019). The relationship between implicit intergroup attitudes and beliefs. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 5862–5871.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual* (Technical Report A-7). University of Florida.

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

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*, 233–248.

Phills, C. E., Kawakami, K., Krusemark, D. R., & Nguyen, J. (2019). Does reducing implicit prejudice increase out-group identification? The downstream consequences of evaluative training on association between the self and social categories. *Social Psychological and Personality Science*, *10*, 26–34.

Phills, C. E., Williams, A., Wolff, J. M., Smith, A., Arnold, R., Felegy, K., & Kuenzig, M. E. (2018). Intersecting race and gender stereotypes: Implications for group-level attitudes. *Group Processes & Intergroup Relations*, *21*, 1172–1184.

Richards, A., & Blanchette, I. (2004). Independent manipulation of emotion in an emotional stroop task using classical conditioning. *Emotion*, *4*, 275–281.

Rudman, L. A., & Goodwin, S. A. (2004). Gender differences in automatic in-group bias: Why do women like women more than men like men? *Journal of Personality and Social Psychology*, *87*, 494–509.

Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, *82*, 919–934.