

# UC Davis

## UC Davis Previously Published Works

### Title

Black + White = Not White: A Minority Bias in Categorizations of Black-White Multiracials

### Permalink

<https://escholarship.org/uc/item/1c15h4s5>

### Authors

Chen, Jacqueline  
Pauker, Kristin  
Gaither, Sarah  
et al.

### Publication Date

2018-01-11

### DOI

10.31234/osf.io/htdea

Peer reviewed

# Reducing Prejudice With Labels: Shared Group Memberships Attenuate Implicit Bias and Expand Implicit Group Boundaries

Personality and Social

Psychology Bulletin

1–11

© 2015 by the Society for Personality

and Social Psychology, Inc

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0146167215621048

pspb.sagepub.com



W. Anthony Scroggins<sup>1</sup>, Diane M. Mackie<sup>1</sup>, Thomas J. Allen<sup>2</sup>,  
and Jeffrey W. Sherman<sup>3</sup>

## Abstract

In three experiments, we used a novel Implicit Association Test procedure to investigate the impact of group memberships on implicit bias and implicit group boundaries. Results from Experiment 1 indicated that categorizing targets using a shared category reduced implicit bias by increasing the extent to which positivity was associated with Blacks. Results from Experiment 2 revealed that shared group membership, but not mere positivity of a group membership, was necessary to reduce implicit bias. Quadruple process model analyses indicated that changes in implicit bias caused by shared group membership are due to changes in the way that targets are evaluated, not to changes in the regulation of evaluative bias. Results from Experiment 3 showed that categorizing Black targets into shared group memberships expanded implicit group boundaries.

## Keywords

implicit bias, shared group memberships, intergroup processes, intervention

Received February 6, 2015; revision accepted November 12, 2015

It appears Shakespeare was wrong—social psychological research has found that a rose by any other name would not smell as sweet. That is because categorization, labeling an entity as a member of a group, fundamentally changes the way we perceive that entity, whether it is a rose or a stink lily (Fiske & Taylor, 1991; Hugenberg & Sacco, 2008; Tajfel & Wilkes, 1963). This phenomenon is well documented with social categories: Changing the social categorization of an individual changes how the individual is both perceived and evaluated. And no other social categorization seems to have the impact that labeling an individual an ingroup versus an outgroup member does: Judgments as fundamental as whether another has a mind or not are affected by whether that other is an ingroup or outgroup member (Hackel, Looser, & Van Bavel, 2014). On the strength of this evidence for the power of categorization, we conducted three experiments to investigate whether and how implicit bias toward outgroup individuals might be reduced if they were recategorized as members of an ingroup.

Social categorization has the well-known effects of making categorized individuals seem more similar both to one another and to the prototype of the target group (for a review, Hugenberg & Sacco, 2008). Perhaps the best experimental evidence for this effect was provided by Levin and Banaji (2006) who assessed perceived skin color as a function of

labeling the same target face as either a Black man or a White man. Participants perceived the stimulus (actually a morphed composite of a Black and White face) labeled *Black* as having darker skin tone than the same stimulus labeled *White*. Social categorization also affects remembered perception. For example, categorizing someone as an uneducated Black or as an assailant leads perceivers to remember the target as having features more stereotypic of Blacks than when the target was categorized as educated or as an embezzler (Benzeev, Dennehy, Goodrich, Kolarik, & Geisler, 2014; Osborne & Davies, 2013; see also Corneille, Huart, Becquart, & Brédart, 2004; Huart, Corneille, & Becquart, 2005).

The social categorization *ingroup member* has been shown to be especially powerful psychologically. Compared with outgroup members, those categorized as ingroup members are viewed as possessing more positive traits (Brewer & Campbell,

<sup>1</sup>University of California, Santa Barbara, USA

<sup>2</sup>Stanford University, CA, USA

<sup>3</sup>University of California, Davis, USA

## Corresponding Author:

W. Anthony Scroggins, Department of Psychological and Brain Sciences,  
University of California, Santa Barbara, UCEN Road, CA 93106, USA.  
Email: wascroggins@gmail.com

1976; Otten & Epstude, 2006) and are evaluated more positively (Mummendey & Schreiber, 1983; Tajfel & Turner, 1986). This is true even when the criterion for distinguishing between ingroups and outgroups is trivial (Tajfel, Billig, Bundy, & Flament, 1971), but the effect is even stronger when the groups are real and meaningful (Mullen, Brown, & Smith, 1992). Based on findings such as these, researchers have tried to combat intergroup bias by making a shared group membership salient, thus conferring the benefits of ingroup membership on former outgroup members.

### Using Shared Group Memberships to Reduce Intergroup Bias

Because most social beings belong to multiple groups, it is likely that people who differ on one categorization, say ethnicity, might share another, say gender or occupation. Making salient such cross-cutting group memberships can reduce bias. For example, any bias that White men feel toward Black men might be reduced by making gender salient, and any bias that White men feel toward White women might be reduced by making ethnicity salient (Crisp, Ensari, Hewstone, & Miller, 2003). However, cross-categorizing groups can redirect prejudice rather than eliminate it. Both ethnicity and gender cross-categorizations make the double outgroup status of Black women even more salient, making evaluations of double outgroups more negative even as it improves evaluations of overlapping groups (Mullen, Migdal, & Hewstone, 2001).

Another approach to recategorizing outgroup members into the ingroup is the common ingroup identity model (Gaertner & Dovidio, 2014; Gaertner, Dovidio, Anastasio, Bachman, & Rust, 1993). According to the common identity model, intergroup bias can be reduced by inducing members of two separate groups to recategorize themselves as members of one inclusive group, and either ignoring their subgroup categories (one-group representation), or simultaneously recognizing their subgroup categories (dual identity representation). Both of these methods change the cognitive representation from an “us” versus “them” orientation, to the more inclusive “we” orientation. Once former outgroup members are included in the shared group, they receive the benefits typically bestowed on ingroup members. In one experiment, for example, Gaertner, Mann, Murrell, and Dovidio (1989) found that intergroup relations were improved when members of two groups were placed at alternating chairs around a table and the larger group was given a new and separate name. Field experiments have also shown that shared group memberships can reduce bias in real-world contexts (Sherif, Harvey, White, Hood, & Sherif, 1961). For example, Nier and colleagues (2001) found that students were more willing to participate in a survey administered by a Black student if the experimenter was wearing clothing that identified them as a supporter of the homeschool, compared with when they wore clothing that identified them as a supporter of

a rival school. In a 3-year longitudinal study, Kessler and Mummendey (2001) found that East Germans who categorized East Germans and West Germans using the shared category “Germans” reported lower levels of explicit bias toward West Germans compared with East Germans who did not categorize West Germans using the “German” shared category. Thus, recategorization in a way that makes “us” and “them” into “we” has been shown to reduce explicit intergroup bias.

Unfortunately, the meaning of a reduction in explicit bias, in general, is not clear. For example, it could reflect increased concern for self-presentation, rather than true attitude change. This may be especially likely in socially sensitive domains. In domains in which self-presentation concerns are low, such as political party preference, explicit measures predict behavior. However, in domains in which self-presentation concerns are high, such as racial preference, explicit measures often do not predict behavior (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). For this reason, we were interested in shared group categories’ ability to reduce *implicit* bias.

### Implicit Racial Bias

Implicit racial bias is a relatively automatic negative evaluation of individuals based on their race. Assessing implicit rather than explicit bias has significant advantages. First, measures of implicit racial bias are less susceptible to social desirability concerns than are measures of explicit racial bias. Second, implicit racial bias may have higher predictive validity than explicit racial bias. Greenwald and colleagues (2009) found that people high in implicit racial bias are more likely to engage in discriminatory behavior. For example, both emergency room and general resident physicians with greater levels of implicit racial bias were less likely to recommend the optimal treatment to Black patients, compared with White patients (Green et al., 2007). Similarly, implicit bias on the part of hiring managers negatively affects minority applicants’ chances of being called back for an interview (Rooth, 2010). Furthermore, police cadets high in implicit racial bias were more likely to shoot unarmed Black targets during a computerized shooting task (Tuttle, 2009).

Given these findings, it is not surprising that there has been a great deal of interest in identifying methods of reducing implicit racial bias, and several studies demonstrate that changes in social categorization affect implicit evaluation. For instance, Black targets were implicitly evaluated more positively when they were dressed as lawyers compared with when they were dressed as prisoners (Barden, Maddux, Petty, & Brewer, 2004). Similarly, Mitchell, Nosek, and Banaji (2003) found that participants exhibited less implicit bias toward specific Black athletes compared with specific White politicians when the targets were evaluated by occupation but *more* implicit bias toward those same Black targets compared with the same White politicians when the targets were evaluated by

race. Other research has shown that listing characteristics shared by both an ingroup and an outgroup can reduce implicit bias (Hall, Crisp, & Suen, 2009). In research most closely relevant to our goals, Van Bavel and Cunningham (2009) showed that White participants assigned to minimal groups that included both Whites and Blacks had more positive implicit evaluations of Black ingroup members than of Black outgroup members. The research described here extends this earlier work by targeting real shared categories (as opposed to minimal groups), using multinomial modeling (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005) to identify the processes by which shared group memberships reduce implicit bias, measuring the effectiveness of using a shared category to reduce the implicit racial bias of Whites and non-Black minorities, and demonstrating that shared categories expand implicit group boundaries.

## Overview

We investigated whether categorizing Black male targets as members of a shared group attenuated non-Black participants' implicit bias toward them. To ensure that implicit bias toward Black targets was not reduced because they were no longer being thought of as Black, we also made the targets' ethnicity salient. In addition, we examined whether making a shared group membership salient would increase the degree to which former outgroup members were implicitly seen as ingroup members. In Experiments 1 and 2, non-Black participants completed Implicit Association Tests (IATs; Greenwald, McGhee, & Schwartz, 1998) that measured their implicit bias toward Black male targets compared with White male targets. In Experiment 1, we compared the degree of implicit bias revealed when male face stimuli were categorized as *Black* versus *White*, compared with when the same stimuli faces were categorized as *Black University of California, Santa Barbara (UCSB) students* (a dual category that includes a shared group membership) compared with *White UCSB students*. We expected that making a shared membership salient would reduce implicit bias toward the targets. Because the IAT is a measure of relative implicit bias, lower levels of implicit bias could be the result of increased positivity toward the target group or decreased positivity toward the comparison group. We predicted that shared membership would reduce implicit bias by increasing the positivity of the Black targets. We developed a novel procedure using the IAT that crossed the manipulated target group category with manipulated comparison categories and analyzed participants' responses with multinomial modeling (Conrey et al., 2005) to identify the source of any changes in implicit bias.

In Experiment 2, we assessed whether including Blacks in the category *UCSB student* reduced implicit racial bias because *UCSB student* is an ingroup category or because it is a category that is viewed positively by UCSB participants. To separate positivity of the label from the inclusiveness of

the label, Experiment 2 compared implicit bias responses to *Blacks* and *Black UCSB students*, and implicit bias responses toward *Black firefighters*, a social category that is as positive as *UCSB student*, but is not a group to which participants belong. Again, subjecting participant responses to multinomial modeling was intended to reveal more about the processes underlying changes in implicit bias. In Experiment 3, we used a modified IAT to investigate whether implicit group boundaries are expanded when shared category labels are used.

## Experiment 1

The purpose of this experiment was twofold: to demonstrate that making a shared category membership salient can reduce implicit racial bias, and to determine whether that reduction in implicit racial bias is due to increased positivity toward Black targets, decreased positivity toward White targets, or both. To test these hypotheses, participants in Experiment 1 took one of four IATs to measure their implicit bias toward Blacks. The labels of the target category (*Black* or *Black UCSB student*) were crossed with the comparison category (*White* or *White UCSB student*).

## Method

**Participants and design.** One hundred forty-six UCSB undergraduate students participated in exchange for course credit. Data from four Black participants were excluded from analysis. All non-Black participants were retained with the assumption that both White and non-Black minority participants would have baseline pro-White/anti-Black implicit attitudes (Blair, Judd, Havranek, & Steiner, 2010; Rudman, Feinberg, & Fairchild, 2002). The final sample of 142 participants (68% women; 39% White, 28% Asian, 25% Latino, 7% "Other") were randomly assigned to one of four conditions in the 2 (target label: *Black* or *Black UCSB student*)  $\times$  2 (comparison label: *White* or *White UCSB student*) between-subjects design.

**Procedure.** On entering the lab, participants were seated in a cubicle with a computer and told that they would be completing a computerized categorization task. They then completed one of four IATs.

**Implicit bias dependent measure.** The IAT is a measure of relative implicit bias. It assesses the strength of association between target categories (e.g., Black person and White person) and attributes (e.g., unpleasant and pleasant). This is achieved by having participants categorize stimuli that belong to one of the four categories (e.g., Black faces or White faces and pleasant words or unpleasant words) by pressing one of two response keys, each of which are assigned to one target category and one attribute. A basic assumption of the IAT is that the more associated two categories are, the easier it is

for participants to correctly categorize stimuli belonging to those two categories when they share a response key. Thus, when highly associated categories share the same response key, categorization latencies are expected to be shorter and responses more accurate compared with when highly associated categories are assigned to different response keys.

During the IAT, participants complete seven different categorization blocks. In these blocks, stimuli appear in the middle of the screen one at a time. The participants must correctly sort the stimulus into the proper category by pressing the appropriate response key. The blocks consist of either 20 or 40 such trials. The blocks of interest are those in which highly associated categories share a response key (compatible block) and those in which highly associated categories share a different response key (incompatible block). The difference in response latencies between these two blocks is used to compute implicit bias.

The stimuli for the four versions of the IAT were identical. The stimuli intended to be categorized as *pleasant* were happy, laughter, love, peace, and pleasure. The stimuli intended to be categorized as *unpleasant* were agony, death, evil, hatred, and sickness. These stimuli words were taken directly from Greenwald and colleagues' (1998) original stimulus set. The attitude objects used were photos of five different individual Black males with neutral expressions and of five different individual White males with neutral expressions. Depending on condition, participants were asked to categorize the Black faces either as *Black* or *Black UCSB student* and the White faces as either *White* or *White UCSB student*.

The IATs followed the procedure outlined by Greenwald et al. (1998). In Block 1 (20 trials), participants categorized faces as either *Black* or *Black UCSB student* by pressing the *e* key and as *White* or *White UCSB student* by pressing the *i* key. In Block 2 (20 trials), participants categorized unpleasant words by pressing the *e* key, and pleasant words by pressing the *i* key. In Block 3 (20 trials), participants practiced using a single response key to categorize both attitude objects and attributes. In this block, participants pressed the *e* key to categorize faces as *Black* or *Black UCSB student* and stimuli as unpleasant and the *i* key to categorize faces as *White* or *White UCSB student* and stimuli as pleasant. Block 4 (40 trials), or the compatible block, was identical to Block 3. In Block 5, participants practiced categorizing *Black* or *Black UCSB student* faces and *White* or *White UCSB student* faces using reversed response keys. That is, *Black* or *Black UCSB student* faces were categorized using the *i* key and *White* or *White UCSB student* faces were categorized using the *e* key. In Block 6 (20 trials), participants again practiced using a single response key to categorize both attitude objects and attributes. In this block, participants pressed the *e* key to categorize *White* or *White UCSB student* faces and unpleasant stimuli and the *i* key to categorize *Black* or *Black UCSB student* faces and pleasant stimuli. Block 7 (40 trials), or the incompatible block, was identical to Block 6. Response

times for all trials were recorded. Differences in mean response times in the compatible block and the incompatible block are used to calculate implicit bias.

After completing the IAT, participants answered demographic questions regarding their age, gender, and ethnicity.<sup>1</sup> They were then debriefed and thanked for their participation.

## Results and Discussion

**Implicit bias.** Response times were converted into IAT *D* scores using the improved scoring algorithm recommended by Greenwald, Nosek, and Banaji (2003). Higher scores indicate more pro-White, anti-Black implicit bias. This resulted in *D* scores for each of the four conditions: *Black/White* ( $M = .69, SD = .34$ ), *Black/White UCSB student* ( $M = .64, SD = .25$ ), *Black UCSB student/White* ( $M = .53, SD = .42$ ), and *Black UCSB student/White UCSB student* ( $M = .46, SD = .36$ ). We hypothesized that when a shared category membership was made salient, this implicit bias would be attenuated, compared with when a shared category membership was not made salient. To test this hypothesis, we conducted a 2 (target category)  $\times$  2 (comparison category) ANOVA. There was a main effect for target category,  $F(1, 138) = 8.74, p = .004$ , 95% confidence interval (CI)<sub>Blk-Blk UCSB</sub> = [.057, .288], partial  $\eta^2 = .06$ ; no main effect for comparison category,  $F(1, 138) = 1.14, p = .288$ ; and no interaction between target category and comparison category,  $F(1, 138) = .01, p = .924$ . These results show that regardless of comparison category, implicit bias toward Black targets was reduced when the shared social category *UCSB student* was made salient ( $M = .49, SD = .38$ ), relative to when it was not ( $M = .66, SD = .29$ ). Even when the comparison category was held constant, implicit bias was reduced when the same faces were categorized as *Black UCSB student* versus *White* ( $D = .46$ ) compared with when faces were categorized as *Black* versus *White* ( $D = .69$ ). To ensure that our manipulations influenced implicit racial bias regardless of participants' ethnicity, we re-analyzed the data including minority status (*White* vs. *Minority*—numbers in some cells did not allow for additional levels of the variable) as a factor in the ANOVA. The 2 (target category)  $\times$  2 (comparison category)  $\times$  2 (minority status) analysis revealed no main effect of minority status,  $F(1, 134) = .14, p = .705$ , nor was there an interaction between target category and minority status,  $F(1, 134) = .50, p = .479$ . These results show that regardless of comparison category, implicit bias toward Black targets was reduced when the shared social category *UCSB student* was made salient ( $M = .49, SD = .38$ ), relative to when it was not ( $M = .66, SD = .29$ ). In addition, the results indicate that the implicit bias scores of non-Black minorities (i.e., Asians, Latinos, and participants who identified as *Other*) did not significantly differ from those of *White* participants, nor were they affected differently by the shared category manipulation. These results suggest that making a shared social category membership salient reduced implicit bias, and that



the reduction in implicit bias was driven more by an increase in positivity toward Black targets labeled *ingroup members* than a decrease in positivity toward White targets. However, this suggestion assumes that the evaluation of comparison categories does not change in relation to changes in target categories, an assumption that multinomial modeling allowed us to verify more definitively (see “Identifying Process With Multinomial Modeling” section).

## Experiment 2

In Experiment 1, we found that categorizing targets into the shared ingroup category *UCSB student* attenuated implicit racial bias (regardless of comparison group). In Experiment 2, we sought to replicate this finding and determine which feature of the shared social category led to reduction in implicit bias. The shared social category used in Experiment 1, *UCSB student*, was both a positive social category and a social category to which the participants belonged. To separate the effects of positivity of the label from effects of inclusiveness of the label, we measured people’s implicit bias toward Blacks versus Whites and compared it with the bias found when identical targets were categorized as either firefighters (positive non-shared) or UCSB students (positive shared). Based on the prominent role ingroup membership plays in intergroup bias (Tajfel & Turner, 1986) and the common ingroup identity model’s tenet that bringing outgroup members into a shared ingroup leads to reductions in intergroup bias (Gaertner & Dovidio, 2014), we hypothesized that the UCSB student label would be more successful at reducing implicit bias toward Blacks than the firefighter label.

## Method

**Participants and design.** One hundred twelve UCSB undergraduate students participated in exchange for course credit. Data from two Black participants were excluded from analysis. Five participants failed to complete the demographics. Their data were included in analyses. The final sample of 110 participants (63% women; 41% White, 26% Asian, 24% Latino, 9% “Other”) was randomly assigned to one of three conditions in the single-factor three-level (IAT: Black/White, Black UCSB student/White UCSB student, Black firefighter/White firefighter) between-participants design.

**Procedure.** Participants took one of three IATs. The face stimuli and the words intended to be positive and negative were the same as those used in Experiment 1. As in Experiment 1, the IATs were identical except for the social category labels used for categorization. In the positive, shared category condition, the target label *Black UCSB student* was paired with the comparison label *White UCSB student*. In the positive, non-shared category condition, the target label *Black firefighter* was paired with the comparison label *White*

*firefighter*. Pre-testing confirmed that ( $N = 39$ ) participants from the same pool rated the categories *firefighter* ( $M = 6.1$ ) and *UCSB student* ( $M = 6.1$ ) *equivalently positively* (where 1 = *very negative* and 7 = *very positive*). In the control condition, the target label *Black* was paired with the comparison label *White*. The procedure for Experiment 2 was identical to the procedure for Experiment 1.

## Results and Discussion

Response times were again converted into IAT  $D$  scores using the recommended improved scoring algorithm. We conducted a one-way between-participants ANOVA to determine whether the three conditions distinctly influenced participants’ level of implicit bias. The ANOVA revealed no effect of condition,  $F(2, 107) = 2.28, p = .107$ . To test our hypothesis that a positive, shared category would reduce implicit bias toward Black targets, relative to the control, to a greater degree than a positive, non-shared category, we ran a pair of planned comparisons. The contrast weights for the first planned comparison were Firefighter (0), Control (1), and UCSB student (−1). The contrast weights for the second planned comparison were Firefighter (−1), Control (1), and UCSB student (0). The planned comparisons revealed that, compared with the control condition ( $M = .58, SD = .31$ ), implicit bias in the positive, shared category condition ( $M = .42, SD = .39$ ) was attenuated,  $t(107) = 1.99, p = .049, 95\% CI_{\text{Blk-Blk UCSB}} = [.00048, .32550], \text{partial } \eta^2 = .036$ . Implicit bias in the positive, non-shared category condition ( $M = .55, SD = .31$ ) was not attenuated, relative to the control,  $t(107) = .34, p = .738$ . Again, to ensure that the levels of implicit racial bias (i.e.,  $D$  scores) of non-Black minority participants were not significantly different from White participants, we conducted another analysis including minority status (White/Minority) as a factor in the ANOVA. This analysis revealed no main effect of minority status,  $F(1, 104) = 1.21, p = .275$ , nor was there an interaction between condition and minority status,  $F(2, 104) = .08, p = .926$ . These results suggest that positivity alone is not sufficient for a social category label to reduce implicit bias. Only when the social category label also represented a shared group, not merely a positive group, was implicit bias reduced.

## Meta-Analyses

As an additional more highly powered test of our hypotheses, and to exclude any influence of the varying comparison conditions in Experiments 1 and 2, we combined the data from the two Black versus White categorization conditions in Experiment 1 and Experiment 2 and the data from the two Black UCSB student versus White UCSB student categorization conditions in Experiments 1 and 2. Implicit bias was significantly reduced in the shared category conditions ( $D = .44$ ) relative to the non-shared category conditions ( $D = .64$ ),  $F(1, 147) = 11.88, p < .001, 95\% CI_{\text{Blk/Wht-Blk UCSB/Wht UCSB}} =$

[.086, .317], partial  $\eta^2 = .075$ . Another meta-analysis combining the three Experiment 1 and Experiment 2 conditions in which participants categorized Black faces as *Black UCSB student* (Black UCSB student vs. White and Black UCSB student vs. White UCSB student in Experiment 1 and Black UCSB student vs. White UCSB student in Experiment 2) revealed significantly lower levels of implicit bias ( $D = .47$ ) relative to the three conditions in the two experiments in which participants categorized Black faces as *Black* (Black vs. White and Black vs. White UCSB student in Experiment 1 and Black vs. White in Experiment 2;  $D = .64$ ),  $F(1, 215) = 13.21$ ,  $p < .001$ , 95% CI<sub>Blk-Blk UCSB</sub> = [.079, .268], partial  $\eta^2 = .058$ . These results provide further support that using a shared ingroup label to categorize Black faces reduced implicit bias.

### Identifying Process With Multinomial Modeling

There are at least three mechanisms that may have produced this effect. First, shared group membership may alter evaluations of outgroup faces by changing the associations activated by those faces. Outgroup faces categorized as belonging to a shared group may activate less negative associations. Second, shared group membership may alter evaluations of ingroup faces by changing the associations activated by those faces. Ingroup faces contrasted against a shared group membership may activate less positive associations. Third, shared group membership may cue perceivers to better control the influence of their associations while responding on the IAT. In this case, evaluations of the targets may be unaffected, but the shared group membership prompts people to inhibit biased responses to the target. All three mechanisms have been shown to be associated with reduced implicit intergroup bias (for reviews, see Calanchini & Sherman, 2013; Sherman et al., 2008), and, of course, all three may operate simultaneously.

To shed light on which of these mechanisms account for our results, we applied the quadruple process model (quad model; Conrey et al., 2005; Sherman et al., 2008) to the IAT data from Experiments 1 and 2. The quad model is a multinomial model designed to estimate the independent contributions of multiple processes to responses on implicit measures of bias. According to the model, performance on implicit measures reflects four distinct processes: activation of associations (AC)—the degree to which associations are activated when encountering a stimulus; detection (D)—the ability to detect the correct response; overcoming bias (OB)—a self-regulatory process that overcomes the influence of associations when they conflict with correct responses; and guessing (G)—general response tendencies that occur when associations are not activated and correct responses cannot be determined.

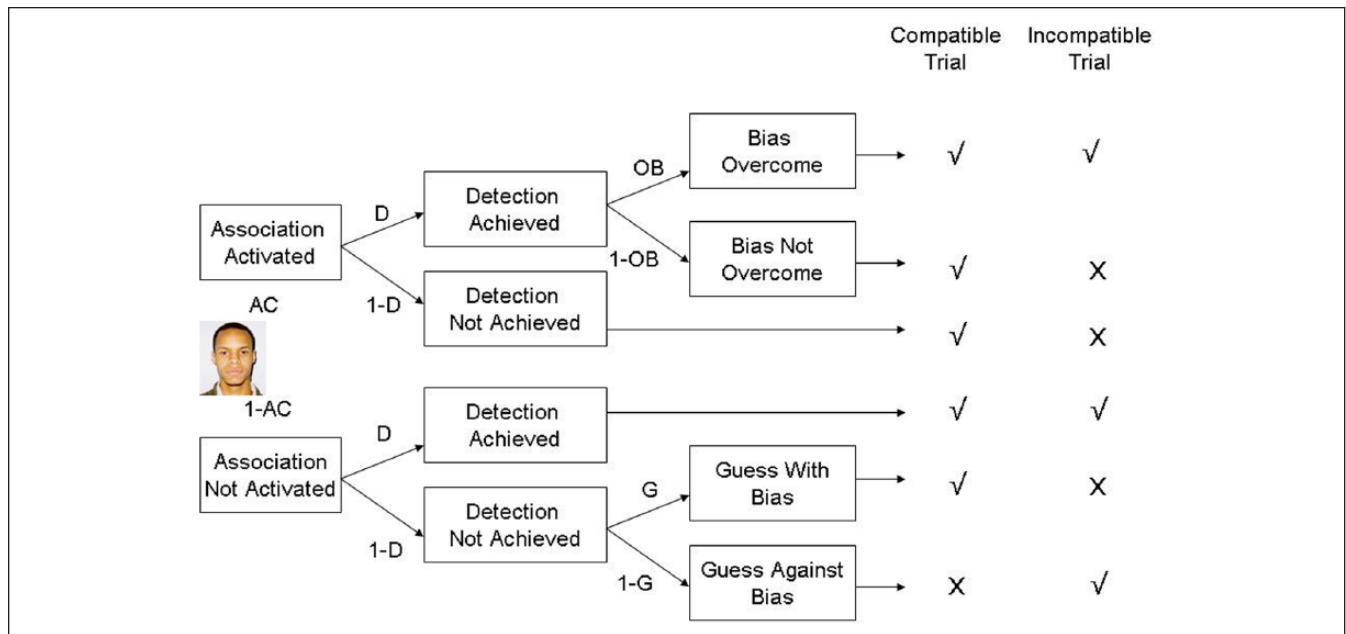
Multinomial models, such as the quad model, attempt to describe experimental outcomes (e.g., error rates) via a set of

variables (or parameters) and a set of equations that establish relationships among the variables. The variables in the equations represent the hypothesized component processes (e.g., activation of associations, detecting a correct response, overcoming bias, guessing). Solving for these variables yields independent estimates of the extent of each process.

The structure of the model is depicted as a processing tree in Figure 1. As an example of how the model works, consider performance on a Black/White evaluative IAT, such as the one used in our experiments. The presentation of a Black face may activate negative associations (AC), predisposing the participant to press the “unpleasant” button. Depending on whether the trial is part of a compatible (Black/unpleasant and White/pleasant) or incompatible (Black/pleasant and White/unpleasant) block, this bias will be either congruent or incongruent with the correct answer “Black” achieved through detection (D). On compatible trials, there is no conflict between activated associations and what is detected. As such, there is no need to overcome the bias (OB) to produce the correct response. However, on incompatible trials, AC and D generate conflicting responses. Which of these two processes ultimately directs the outcome is determined by whether or not the participant succeeds in overcoming his or her bias. Finally, if no association is activated and the correct response cannot be ascertained, then participants must guess (G).

Note that the relationships among the parameters do not imply a serial order in the onset and conclusion of the different processes. Rather, these relationships are mathematical descriptions of the manner in which the parameters interact to produce behavior. Thus, the activation of associations (AC), attempts to detect a correct response (D), and attempts to overcome associations (OB) may occur simultaneously. However, in determining a response on an incompatible trial, the status of OB determines whether AC or D drives responses.

The relationships described by the model form a system of equations that predict the number of correct and incorrect responses for different item types (e.g., Black faces; White faces; pleasant words; unpleasant words) in different conditions (e.g., compatible and incompatible trials). For example, a Black face on an incompatible trial will be responded to correctly with the following probability:  $AC \times D \times OB + (1 - AC) \times D + (1 - AC) \times (1 - D) \times (1 - G)$ . This equation sums the three possible paths by which a correct answer can be returned in this case. The first part of the equation,  $AC \times D \times OB$ , is the likelihood that the association between Black and unpleasant is activated and that the correct answer can be detected and that the association is overcome in favor of the detected response. The second part of the equation,  $(1 - AC) \times D$ , is the likelihood that the association is not activated and that the correct response can be detected. Finally,  $(1 - AC) \times (1 - D) \times (1 - G)$ , is the likelihood that the association is not activated and the correct answer cannot be detected and that the participant guesses correctly.



**Figure 1.** The quadruple process model.

Note. Each path represents a likelihood. Parameters with lines leading to them are conditional on all preceding parameters. The table on the right side of the figure depicts correct (✓) and incorrect (X) responses as a function of process pattern. AC = activation of associations; D = detection; OB = overcoming bias; G = guessing.

The respective equations for each item category (e.g., Black faces, White faces, pleasant words, and unpleasant words in both compatible and incompatible blocks) are then used to predict the observed proportions of errors in a given data set. The model’s predictions are then compared with the actual data to determine the model’s ability to account for the data. A chi-square estimate is computed for the difference between the predicted and observed errors. To best approximate the model to the data, the parameter values are changed through maximum likelihood estimation until they produce a minimum possible value of the chi-square. The final parameter values that result from this process are interpreted as relative levels of the processes. The quad model and the construct validity of its parameters have been extensively validated in previous research (Conrey et al., 2005; Sherman et al., 2008). To identify the processes responsible for our findings, we calculated quad model parameter estimates as a function of condition. For each condition, we estimated two AC, one D, one OB, and one G parameter. The G parameter was coded so that higher scores represented a bias toward guessing with the pleasant key. Because the IAT is a measure of relative evaluations of two groups, estimating the independent evaluative associations for each group is impossible in the standard reaction time analysis. An advantage of the quad model is that it can produce separate estimates for different types of associations. As such, we created one AC parameter that measured the extent to which associations between “Black” and “unpleasant” were activated in performing the task and another AC parameter that measured

the extent to which associations between “White” and “pleasant” were activated in performing the task. We modeled two different sets of data, mirroring the meta-analyses previously described. For both the standard Black categorization conditions and for Black UCSB, we estimated one AC parameter that measured the association between Black faces and unpleasant words, one AC parameter that measured the association between White faces and pleasant words, one OB, one D, and one G parameter. If the decreases in bias seen in shared category contexts were due to a change in the evaluation of the outgroup stimuli, then we would expect lower levels of Black-unpleasant AC in the shared group membership condition than in the standard outgroup condition. In contrast, if the decreases in bias seen in shared category contexts were due to a change in the evaluation of the ingroup stimuli, then we would expect lower levels of White-pleasant AC in the shared group membership condition than in the standard outgroup condition. Finally, if the effects were due to self-regulation, then we would expect higher levels of OB in the shared group membership condition than in the standard outgroup condition.

First, we combined the data from the two Black versus White conditions in Experiments 1 and 2 and compared them with the combined data from the two Black UCSB versus White UCSB conditions in Experiments 1 and 2. The chi-square for model fit was 20.96,  $p = .002$ . Chi-square tests are dependent on sample size, such that minute deviations from the model can jeopardize model fit when power is high (see Cohen, 1988). However, the effect size of lack of model fit



between the actual data and the model's predicted data was small,  $w = .03$ , indicating satisfactory fit when controlling for power. Results showed that the AC Black-bad parameter was significantly lower in the shared ingroup conditions than in the standard Black versus White condition,  $G^2(1) = 9.49, p = .002$ . The AC White-good parameter did not differ between conditions,  $G^2(1) = .06, p = .81$ . The OB parameter also did not differ between the conditions,  $G^2(1) = 1.62, p = .20$ . The D parameter was marginally higher in the shared ingroup than the standard condition,  $G^2(1) = 3.09, p = .08$ . The G parameter did not differ between the conditions,  $G^2(1) = 1.95, p = .16$ .

For the second analysis, we combined the data from all three conditions in Experiments 1 and 2 in which participants categorized faces as Black and compared them with the combined data from all three of the conditions from the two experiments in which participants categorized the targets as Black UCSB students. The chi-square for model fit was 21.39,  $p = .002$ , and, again, the effect size of lack of model fit between the actual data and the model's predicted data was small,  $w = .03$ , indicating satisfactory fit when controlling for power. Results showed that the AC Black-bad parameter was again significantly lower in the shared ingroup conditions than in the standard Black condition,  $G^2(1) = 3.96, p = .047$ . The AC White-good parameter did not differ between conditions,  $G^2(1) = .03, p = .86$ . The OB parameter also did not differ between the conditions,  $G^2(1) = 0.26, p = .61$ . The D parameter was higher in the shared group than the standard outgroup condition,  $G^2(1) = 9.09, p = .003$ . The G parameter did not differ between the conditions,  $G^2(1) = 0.60, p = .44$ .

In sum, the results of the quad model analyses indicate that changes in implicit bias caused by shared group membership are related to changes in the way that outgroup targets are evaluated, not to changes in the evaluation of ingroup targets or to the regulation of evaluative bias. When the same Black faces are categorized as members of shared ingroups, the associations activated are less negative. We found no evidence that shared group membership altered evaluations of ingroup targets or that it encouraged inhibition of biased responses. Interestingly, we did find that shared group membership is associated with increased Detection of the correct responses on the task. Consistent with work on the own-race bias in face perception (Hugenberg, Young, Bernstein, & Sacco, 2010), this suggests that targets receive increased scrutiny to the extent that they are perceived as belonging to an ingroup.

### Experiment 3

Experiments 1 and 2 showed that categorizing targets using a shared group membership can reduce implicit racial bias, and that the reduction was related to the target category becoming more positive when associated with the shared category. Moreover, this reduction in implicit bias was not due to the mere positivity of the shared group membership,

less positive evaluations of ingroup members, or to greater regulation of evaluative bias. Together, the results of these two experiments suggest that inclusion in a shared category was what conferred positivity on the outgroup targets and thus reduced bias. In Experiment 3, we confirmed the operation of this mechanism by testing the idea that making a shared ingroup membership salient increased the degree to which Black targets are seen as ingroup members. The goal of this experiment was to determine for the first time whether shared social category labels expand implicit group boundaries: Are Black UCSB students seen more as ingroup members than Blacks are? We used a modified IAT to measure the degree to which Black or Black UCSB targets and White or White UCSB targets are associated with the labels *ingroup* and *outgroup*.

### Method

**Participants and design.** One hundred fifty-one UCSB undergraduate students participated in exchange for a US\$5 gift card. Data from four Black participants were excluded from analysis. An additional six participants were excluded for failing to follow instructions or due to computer error. The final sample of 141 participants (70% women; 40% White, 33% Asian, 19% Latino, 8% "Other") were randomly assigned to one of four conditions in the 2 (target label: Black or Black UCSB student)  $\times$  2 (comparison label: White or White UCSB student) between-participants design.

**Procedure.** Participants took one of four IATs. As in Experiment 1, the IATs were identical except for the social category labels used for categorization. The target labels were either *Black* or *Black UCSB student*, whereas the comparison labels were either *White* or *White UCSB student*. The same Black and White face stimuli used in Experiments 1 and 2 were used again here. The stimuli intended to be categorized as *ingroup* were five distinct UCSB logos. The stimuli intended to be categorized as *outgroup* were five logos from universities around the country: University of Alabama, University of Oregon, Ohio State University, University of Texas, and the University of Michigan. These IATs were thus intended to assess implicit associations of the Black and White faces with the ingroup or outgroup as a function of how they were categorized. The procedure for Experiment 3 was again identical to the procedure for Experiment 1.

### Results and Discussion

Again, response times were converted into IAT *D* scores. This resulted in *D* scores for each of the four conditions: Black/White ( $M = .44, SD = .39$ ), Black/White UCSB student ( $M = .48, SD = .42$ ), Black UCSB student/White ( $M = .24, SD = .46$ ), and Black UCSB student/White UCSB student ( $M = .39, SD = .41$ ). To determine whether making a shared category membership salient expands implicit group

boundaries, we conducted a 2 (target category)  $\times$  2 (comparison category) ANOVA. There was a main effect for target category,  $F(1, 137) = 4.11, p = .045, 95\% CI_{\text{Blk-Blk UCSB}} = [.004, .283], \text{partial } \eta^2 = .029$ . These results show that Black targets were implicitly seen more as ingroup members when they were categorized using the UCSB student label ( $M = .31, SD = .43$ ), relative to when they were not ( $M = 0.46, SD = .40$ ). There was no main effect for comparison category,  $F(1, 137) = 1.57, p = .213$ , and no interaction between target category and comparison category,  $F(1, 137) = .62, p = .431$ . To ensure that the implicit group boundaries of non-Black minority participants were not significantly different from White participants, we again included minority status (White/Minority) as a factor in the ANOVA, which revealed no main effect of minority status,  $F(1, 133) = .64, p = .426$ , and no interaction between target category and minority status,  $F(1, 133) = .26, p = .613$ . These results indicate that the implicit group boundaries of non-Black minorities (i.e., Asians, Latinos, and participants who identified as Other) did not differ significantly from those of White participants, both in the control condition and as a result of the shared category manipulation. These findings indicate that categorizing Black faces using an ingroup led to an expansion of implicit group boundaries.

## General Discussion

This research provides the first evidence of how and why categorization in terms of shared group memberships can reduce implicit bias. In Experiment 1, categorizing Black targets in terms of the shared group *UCSB student* reduced implicit bias compared with when targets were categorized only as Black. Importantly, however, the results from Experiment 2 showed that categorizing an outgroup member as a member of a universally agreed-on positive group (firefighters) was not sufficient to reduce implicit bias; bias was only reduced when Blacks were categorized as a member of a shared ingroup. Experiment 3 confirmed that, as intended, categorizing targets in terms of an ingroup membership increased the degree to which targets were implicitly seen as ingroup members. These findings applied to both White and non-Black minority participants. Non-Black minorities showed at least as much pro-White/anti-Black bias as White participants, and the two groups of participants were equally affected by the shared category manipulation. Neither did the implicit group boundaries of non-Black minorities differ significantly from those of White participants.

Application of the quad model to the responses in Experiments 1 and 2 indicated that once such ingroup belonging was conferred, bias was reduced, and this reduction was associated with increased positive evaluations of Black faces compared with when those same faces were not included in the ingroup. Interestingly, we found no evidence that the reduction in bias associated with ingroup membership was dependent on overcoming of biased responses. This

makes shared ingroup membership a particularly appealing practical approach to reducing bias, as positive associations seemed to be conferred as part of ingroup membership, rather than as a result of other additional processes.

These studies leave some aspects of our interpretations of our findings unresolved. Implicit bias and implicit group boundaries were not both measured in a single study. This limitation prohibits us from determining whether expansion of group boundaries was causally responsible for the reduction in implicit bias in Experiments 1 and 2. Furthermore, we did not measure explicit bias, so we are unable to investigate the relationships among implicit bias, implicit group boundaries, and explicit bias. Last, we did not investigate the impact of a negative shared category (e.g., polluter). Without this condition, we are unable to determine whether a positive shared category is necessary to reduce implicit bias. It is possible that simply sharing a group membership, whether positive or negative, would be enough to expand implicit group boundaries and reduce implicit bias.

Despite these unresolved issues, this research raises interesting questions that warrant future research. As previously mentioned, further research is necessary to show whether the reduction in implicit bias found in Experiments 1 and 2 was mediated by the expansion of implicit boundaries found in Experiment 3. If all the various methodological and psychological issues associated with administering two similar sequential IATs could be addressed, such evidence could presumably be found by administering an ingroup boundary IAT followed by a racial bias IAT, crossing all possible combinations of *Black* versus *White* and *Black UCSB* versus *White UCSB* labels (as was done in Experiment 1). Another option for testing the relation between implicit ingroup boundaries and implicit racial bias would be to have participants complete either a Black/White IAT or a Black UCSB student/White UCSB student IAT followed by a priming task. During the priming task, participants might be exposed to a series of White and Black male faces. After each exposure, participants would be presented with a UCSB logo or the logo of another university and asked to categorize the logo as either *Us* (for UCSB logos) or *Them* (for non-UCSB university logos) as quickly as possible. If categorizing Black targets as UCSB students expands group boundaries, then we would expect there to be a smaller difference in latencies when categorizing UCSB logos as *Us* following Black and White primes after taking the Black UCSB student/White UCSB student IAT, compared with after taking the Black/White IAT.

Additional research is also necessary to show whether the beneficial effects of categorizing individual outgroup members using an ingroup membership generalizes to other members of that outgroup. This question has been investigated with regard to explicit attitudes (e.g., Brown & Hewstone, 2005; González & Brown, 2006), but research in the implicit attitudes domain is lacking. For instance, would categorizing Black targets as *UCSB students* in a first IAT reduce implicit

bias in a second IAT in which Black targets were categorized as *Black*, compared with administering IATs in the opposite order? Furthermore, would using a shared category label reduce prejudiced behavior? For example, would UCSB participants evaluate the job application of a male who identified himself as a “Black UCSB student” more positively than the application of a male who identified himself as “Black”?

## Conclusion

One of the major benefits of this research is that it readily lends itself to the development of interventions. Whereas it might be difficult to have (especially prejudiced) people think of admired Black Americans and disliked White Americans before cross-ethnic interactions, as other techniques shown to reduce implicit racial bias require (Dasgupta & Greenwald, 2001), it is relatively easy to make shared group memberships salient (because Blacks and non-Blacks share age, gender, occupational, political affiliation, college, and multiple other social category memberships). Based on previous research demonstrating the pernicious consequences of implicit racial bias, it seems likely that such interventions reducing implicit racial bias might improve the level of medical care that Black patients receive (Green et al., 2007), increase the chances that hiring managers would call qualified Black job candidates in for interviews (Rooth, 2010), and reduce the likelihood of unarmed Black men being shot by police officers (Tuttle, 2009).

## Acknowledgment

We thank Melissa Buckley, Mina Chen, Megan Farrell, Yasmin Irfani, and Isella Vite-Arguello for their assistance with data collection.

## 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: This research was supported by the National Science Foundation Graduate Research Fellowship Grant (DGE 1144085) awarded to W. Anthony Scroggins.

## Note

1. Participants also completed an exploratory University of California, Santa Barbara (UCSB) identification scale. The validity of the scale has not yet been determined and, thus, we are not comfortable using it as a basis for claims regarding the relationship between implicit racial bias and UCSB student identification. As a result, it will not be discussed further.

## Supplemental Material

The online supplemental material is available at <http://pspb.sagepub.com/supplemental>.

## References

- Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology, 87*, 5-22.
- Ben-Zeev, A., Dennehy, T. C., Goodrich, R. I., Kolarik, B. S., & Geisler, M. W. (2014). When an “educated” Black man becomes lighter in the mind’s eye: Evidence for a skin tone memory bias. *SAGE Open, 4*, 1-9.
- Blair, I. V., Judd, C. M., Havranek, E. P., & Steiner, J. F. (2010). Using community data to test the discriminant validity of ethnic/racial group IATs. *Zeitschrift für Psychologie/Journal of Psychology, 218*, 36-43.
- Brewer, M. B., & Campbell, D. T. (1976). *Ethnocentrism and intergroup attitudes: East African evidence*. New York, NY: Sage Publications.
- Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. *Advances in Experimental Social Psychology, 37*, 255-343.
- Calanchini, J., & Sherman, J. W. (2013). Implicit attitudes reflect associative, non-associative, and non-attitudinal processes. *Social & Personality Psychology Compass, 7*, 654-667.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology, 89*, 469-487.
- Corneille, O., Huart, J., Becquart, E., & Brédart, S. (2004). When memory shifts toward more typical category exemplars: Accentuation effects in the recollection of ethnically ambiguous faces. *Journal of Personality and Social Psychology, 86*, 236-250.
- Crisp, R., Ensari, N., Hewstone, M., & Miller, N. (2003). A dual-route model of crossed categorisation effects. *European Review of Social Psychology, 13*, 35-73.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81*, 800-814.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York, NY: McGraw-Hill Book.
- Gaertner, S. L., & Dovidio, J. F. (2014). *Reducing intergroup bias: The common ingroup identity model*. Philadelphia, PA: Psychology Press.
- Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., & Rust, M. C. (1993). The common ingroup identity model: Recategorization and the reduction of intergroup bias. *European Review of Social Psychology, 4*, 1-26.
- Gaertner, S. L., Mann, J., Murrell, A., & Dovidio, J. F. (1989). Reducing intergroup bias: The benefits of recategorization. *Journal of Personality and Social Psychology, 57*, 239-249.
- González, R., & Brown, R. (2006). Dual identities in intergroup contact: Group status and size moderate the generalization of positive attitude change. *Journal of Experimental Social Psychology, 42*, 753-767.
- Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis deci-



- sions for black and white patients. *Journal of General Internal Medicine*, 22, 1231-1238.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197-216.
- 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.
- Hackel, L. M., Looser, C. E., & Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: The role of social identity, collective identification, and intergroup threat. *Journal of Experimental Social Psychology*, 52, 15-23.
- Hall, N. R., Crisp, R. J., & Suen, M. W. (2009). Reducing implicit prejudice by blurring intergroup boundaries. *Basic and Applied Social Psychology*, 31, 244-254.
- Huart, J., Corneille, O., & Becquart, E. (2005). Face-based categorization, context-based categorization, and distortions in the recollection of gender ambiguous faces. *Journal of Experimental Social Psychology*, 41, 598-608.
- Hugenberg, K., & Sacco, D. F. (2008). Social categorization and stereotyping: How social categorization biases person perception and face memory. *Social & Personality Psychology Compass*, 2, 1052-1072.
- Hugenberg, K., Young, S., Bernstein, M., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the cross race recognition deficit. *Psychological Review*, 117, 1168-1187.
- Kessler, T., & Mummendey, A. (2001). Is there any scapegoat around? Determinants of intergroup conflicts at different categorization levels. *Journal of Personality and Social Psychology*, 81, 1090-1102.
- Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General*, 135, 501-512.
- Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, 132, 455-469.
- Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, 22, 103-122.
- Mullen, B., Migdal, M. J., & Hewstone, M. (2001). Crossed categorization versus simple categorization and intergroup evaluations: A meta-analysis. *European Journal of Social Psychology*, 31, 721-736.
- Mummendey, A., & Schreiber, H. J. (1983). Better or just different? Positive social identity by discrimination against, or by differentiation from outgroups. *European Journal of Social Psychology*, 13, 389-397.
- Nier, J. A., Gaertner, S. L., Dovidio, J. F., Banker, B. S., Ward, C. M., & Rust, M. C. (2001). Changing interracial evaluations and behavior: The effects of a common group identity. *Group Processes & Intergroup Relations*, 4, 299-316.
- Osborne, D., & Davies, P. G. (2013). Eyewitness identifications are affected by stereotypes about a suspect's level of perceived stereotypicality. *Group Processes & Intergroup Relations*, 16, 488-504.
- Otten, S., & Epstude, K. (2006). Overlapping mental representations of self, ingroup, and outgroup: Unraveling self-stereotyping and self-anchoring. *Personality and Social Psychology Bulletin*, 32, 957-969.
- Rooth, D. O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, 17, 523-534.
- Rudman, L. A., Feinberg, J., & Fairchild, K. (2002). Minority members' implicit attitudes: Automatic ingroup bias as a function of group status. *Social Cognition*, 20, 294-320.
- Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. (1961). *Intergroup conflict and cooperation: The Robbers Cave experiment*. Norman, OK: University Book Exchange.
- Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, 115, 314-335.
- Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology*, 1, 149-178.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7-24). Chicago, IL: Nelson-Hall.
- Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgment. *British Journal of Psychology*, 54, 101-114.
- Tuttle, K. M. K. (2009). *Implicit racial attitudes and law enforcement shooting decisions*. Unpublished manuscript. University of Michigan.
- Van Bavel, J. J., & Cunningham, W. A. (2009). Self-categorization with a novel mixed-race group moderates automatic social and racial biases. *Personality and Social Psychology Bulletin*, 35, 321-335.