Title
Mechanisms of Group Membership and Exemplar Exposure Effects on Implicit Attitudes

Journal
Social Psychology, 41(3)

ISSN
1864-9335

Authors
Gonsalkorale, Karen
Allen, Thomas J
Sherman, Jeffrey W
et al.

Publication Date
2010

DOI
10.1027/1864-9335/a000023

Peer reviewed

Original Article

Mechanisms of Group Membership and Exemplar Exposure Effects on Implicit Attitudes

Karen Gonsalkorale[1], Thomas J. Allen[2], Jeffrey W. Sherman[2], and Karl Christoph Klauer[3]

[1]University of Sydney, Australia, [2]University of California, Davis, CA, USA, [3]Universität Freiburg, Germany

Abstract. This research examines the mechanisms underlying group-based differences in implicit attitudes and malleability of implicit attitudes resulting from exposure to exemplars. We tested whether these effects are due to differences in activated associations or to the regulation of those associations. In Study 1, Black participants exhibited less pro-White bias and activation of pro-White and anti-Black associations compared with White participants. In Study 2, exposure to positive Black and negative White exemplars produced lower pro-White bias and less activation of biased associations. These findings indicate that group-based differences in implicit attitudes and exemplar exposure effects reflect variability in and malleability of automatic associations. Implications for understanding the role of contact on intergroup attitudes are discussed.

Keywords: implicit prejudice, ingroup bias, attitude change, automatic associations, self-regulation

When implicit attitude measures first burst onto the social psychology scene, they were often conceived of as uncontaminated reflections of automatic associations resistant to change. An accumulated body of evidence has since shown that scores on implicit attitude measures are responsive to variations among respondents, stimuli, and context (see Blair, 2002; Sherman, Gawronski, Gonsalkorale, Hugenberg, Allen, & Groom, 2008). It is now clear that implicit task performance varies across individuals and different situations. However, though this literature suggests that implicit attitudes are variable and malleable, relatively little is known about the mechanisms underlying these effects. Exactly what do these findings mean? In this paper, we address this question by testing competing explanations for two well-documented and related effects in implicit attitudes research: (1) group-based differences in implicit attitudes and (2) malleability of implicit attitudes resulting from exposure to group exemplars.[1]

Research has shown that implicit attitudes vary as a function of the perceiver's group membership. The robust finding of ingroup bias on explicit measures (Hewstone, Rubin, & Willis, 2002) is often reflected on implicit measures as well (e.g., Ashburn-Nardo, Voils, & Monteith, 2001; Fazio, Jackson, Dunton, & Williams, 1995; Greenwald, McGhee, & Schwartz, 1998; Jellison, McConnell, & Gabriel, 2004; Olson, Crawford, & Devlin, 2009; Perdue, Dovidio, Gurtman, & Tyler, 1990; Rudman, Greenwald, Mellott, & Schwartz, 1999). For instance, whereas Korean Americans show implicit preference for Koreans over Japanese, the reverse is true for Japanese Americans (Greenwald et al., 1998). Similarly, White Americans on average display a robust preference for White over Black on implicit measures (Nosek, Banaji, & Greenwald, 2002). However, this pattern of implicit ingroup favoritism is not always evident. Of most direct relevance to the current research, African Americans have been shown to display implicit ingroup preference (Fazio et al., 1995; Olson et al., 2009), no evidence of bias (Jost, Banaji, & Nosek, 2004; Nosek et al., 2002), or weak levels of pro-White bias (Ashburn-Nardo, Knowles, & Monteith, 2003).

One important determinant of ingroup favoritism is the extent of contact with outgroup members. There is an abundance of evidence that increased contact with outgroup members reduces expressions of prejudice on explicit measures (Pettigrew & Tropp, 2006). Likewise, growing evidence suggests that intergroup contact reduces implicit prejudice. Self-report-

[1] We use the term "implicit attitude" to refer simply to an attitude that is measured with an implicit measure. Though "implicit attitudes" are defined as the behavioral outcomes of implicit measures (e.g., reaction time effects), they are not assumed to be isomorphic with the underlying evaluative associations that instigate responses on the measure. Rather, behavioral biases on implicit measures (i.e., "implicit attitudes") may or may not correspond closely with underlying associations, depending on the intervention of other processes that translate those associations into behavioral responses on the implicit measures. Thus, we call the behavioral bias an "implicit attitude" in the common vernacular, but distinguish this "attitude" from evaluative associations.

DOI: 10.1027/1864-9335/a000023

ed contact with members of an outgroup predicts less implicit prejudice toward that group (Dasgupta & Rivera, 2008; Tam, Hewstone, Harwood, Voci, & Kenworthy, 2006), particularly when the contact is with close others (Aberson, Shoemaker, & Tomolillo, 2004; Lemm, 2006). Conversely, implicit ingroup favoritism is more likely to occur among individuals who have higher proportions of ingroup (versus outgroup) family members and acquaintances (Greenwald et al., 1998) or who had greater positive contact with ingroup members during childhood (Ashburn-Nardo, Monteith, Arthur, & Bain, 2007).

Mirroring the work on intergroup contact, other research found that attitudes shift in response to exposure to group exemplars. Brief exposure to popular Black and disliked White exemplars reduces pro-White bias on both explicit and implicit measures, relative to exposure to popular Whites and disliked Blacks or to unfamiliar racial exemplars (Bodenhausen, Schwarz, Bless, & Wanke, 1995; Dasgupta & Greenwald, 2001; Govan & Williams, 2004; Mitchell, Nosek, & Banaji, 2003). Exposure to counterstereotypic exemplars through mental imagery has been shown to produce the same effects in the domain of gender stereotyping (Blair, Ma, & Lenton, 2001). Furthermore, White individuals show less implicit bias in the presence of a Black experimenter than a White experimenter (Lowery, Hardin, & Sinclair, 2001) and when they anticipate contact with a Black partner assigned to a high-power versus low-power role (Richeson & Ambady, 2003). Thus, both live contact with and virtual exposure to positive or counterstereotypic exemplars reduces the extent of implicit bias.

Although group-based differences in implicit bias and exemplar exposure effects on implicit bias have been well documented, their underlying mechanisms are not well understood. The present research seeks to specify those mechanisms and determine if group membership effects and exemplar exposure effects share the same mechanisms. Dasgupta and Rivera (2008) argued that frequent positive contact with group members influences the associations that are brought to mind automatically by making positive exemplars chronically accessible. These associations may account both for the general finding of implicit ingroup bias (due to more frequent positive contact with ingroup versus outgroup members) as well as the reduced implicit bias among individuals experiencing high levels of intergroup contact. Similarly, Dasgupta and Greenwald (2001) proposed that brief exposure to counterattitudinal exemplars increases activation of positive associations with the group, resulting in less biased implicit evaluations (also see Blair et al., 2001). Thus, one possibility is that group-based differences in implicit bias and exemplar exposure effects on implicit bias both result from the activation of more or less favorable group associations.

However, a competing account suggests that these effects may be due to differences in self-regulatory processes rather than automatic associations. Despite having different contact experiences, majority and minority members live in a society that is dominated by the majority culture. Researchers have proposed that promajority associations at the societal level seep into the minds of minority individuals (Karpinski & Hilton, 2001; Olson & Fazio, 2004). This suggests that members of disadvantaged groups may have similar associations to members of advantaged groups. Given that people are motivated to view their social groups favorably (Tajfel & Turner, 1979), minority members may be particularly likely to engage in self-regulation to overcome activated pro-majority associations. This would lead minority members to exhibit lower levels of implicit bias compared to majority members, despite having equally strong pro-majority associations.

Self-regulation of activated associations may likewise be responsible for exemplar exposure effects on implicit bias. Intergroup contexts trigger concerns about appearing prejudiced among majority members (Shelton, 2003), leading them to engage in self-control when interacting with or perceiving outgroup members (Richeson & Shelton, 2003; Richeson et al., 2003; Richeson, Trawalter, & Shelton, 2005; Shelton, 2003). Moreover, research suggests that self-regulatory effort increases when the potential for bias is made salient directly, via manipulation of instructions (Shelton, 2003), or indirectly, via the manipulation of stimulus features (Maddux, Barden, Brewer, & Petty, 2005). With regard to stimulus features, Maddux et al. (2005) proposed that individuals engage in greater self-regulation when they are exposed to social contexts that signal the potential for a prejudiced response (Maddux et al., 2005). This view is supported by recent evidence that the placement of outgroup members in positive contexts reduces implicit bias via the engagement of self-regulatory processes that override activated associations (Allen, Sherman, & Klauer, in press). Exposure to positive outgroup exemplars could similarly highlight the potential for prejudiced responses and increase self-regulatory effort. This suggests that exemplar exposure effects could result not from differences in activation of automatic associations, but from differences in the extent to which those associations are regulated.

To analyze the mechanisms underlying group-based differences in implicit attitudes and exemplar exposure effects, the contributions of automatic associations and self-regulation need to be quantified independently. Previous studies have assessed automatic associations using implicit attitude measures, but these measures do not tap variability in automatic associations alone. A growing body of research has shown that performance on implicit measures reflects multiple processes, both automatic and controlled (Amodio, Harmon-Jones, Devine, Curtin, Hartley, & Covert, 2004; Bartholow, Dickter, & Sestir, 2006; Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Payne, 2001; Sherman, 2009; Sherman et al., 2008). Of most direct relevance to the current research, responses on implicit measures have been shown to be affected by participants' ability to regulate the influence of automatic associations during completion of the measures (Allen et al., in press; Amodio et al., 2004; Bartholow et al., 2006; Conrey et al., 2005; Gonsalkorale, Sherman, & Klauer, 2009; Gonsalkorale, von Hippel, Sherman, & Klauer, 2009; Sherman et al., 2008). Thus, scores on implicit measures reflect both the strength of automatic associations and people's ability to overcome those associations when completing the measure.

One method to separate the strength of automatic associations from the ability to overcome them is the Quadruple Process model (Quad model; Sherman et al., 2008). The Quad model is a multinomial model (see Batchelder & Riefer, 1999) designed to estimate the independent contributions of multiple processes from responses on implicit measures of bias (for reviews of this approach, see Sherman, 2006; Sherman et al., 2008). According to the model, responses on implicit measures of bias reflect the operation of four qualitatively distinct processes: activation of associations (AC), detection (D), overcoming Bias (OB), and guessing (G). The AC parameter refers to the degree to which biased associations (e.g., between Black people and negativity) are automatically activated when responding to a stimulus. All else being equal, the stronger the associations, the more likely they are to be activated and to influence behavior. The D parameter reflects a relatively controlled process that discriminates between contextually appropriate and inappropriate responses. For example, on incompatible trials of the Implicit Association Test (IAT; Greenwald et al., 1998), the appropriate response to a Black face stimulus is to press the "Black or pleasant" key. However, activated associations between Black people and negativity will conflict with the detected correct response. In such cases, the Quad model proposes that an overcoming bias process resolves the conflict. As such, the OB parameter refers to self-regulatory efforts that prevent automatically activated associations from influencing behavior when they conflict with detected correct responses. Finally, the G parameter reflects general response tendencies that may occur when individuals have no associations that direct behavior, and they are unable to detect the appropriate response. Guessing can be random, but it may also reflect a systematic tendency to prefer a particular response. For example, incorrectly categorizing a target face stimulus as "unpleasant" in the IAT could be considered a socially undesirable response. To avoid that possibility, participants may adopt a conscious guessing strategy to respond with the positive rather than the negative key. In other cases, the G parameter may reflect more unconscious biases, such as the tendency to respond with the dominant hand. The Quad model and the construct validity of its parameters have been extensively validated in previous research (see Beer, Stallen, Lombardo, Gonsalkorale, Cunningham, & Sherman, 2008; Conrey et al., 2005; Gonsalkorale, Sherman et al., 2009; Gonsalkorale , von Hippel et al., 2009; Sherman et al., 2008). We used the Quad model in the current research to test whether group-based differences in implicit attitudes (Study 1) and exemplar exposure effects (Study 2) are due to differences in automatic associations, self-regulation, or a combination of these processes.

Study 1

This study examines the processes underlying group-based differences in implicit attitudes. Previous research showed that, on average, White participants display a pro-White bias on the IAT, whereas Black participants show no bias (e.g., Nosek et al., 2002; Stewart, von Hippel, & Radvansky, 2009).[2] To test the competing explanations for this effect, we analyzed data collected from Black and White respondents to the IAT demonstration website (http://implicit.harvard.edu/). If differences in pro-White bias among White and Black participants are due to the different associations activated among the participants, then Black participants should show lower levels of pro-White and anti-Black association activation than White participants. Alternatively, Black and White individuals may instead differ in their level of self-regulation (Allen et al., in press; Maddux et al., 2005), which would be reflected as greater levels of overcoming bias among Black participants.

Method

Participants

Participants were 2,232 Black and 16,456 White individuals selected randomly from among a group of approximately 300,000 respondents who visited the IAT demonstration website (http://implicit.harvard.edu/; Nosek et al., 2002) between December 2002 and May 2006.

Procedure

After providing demographic information, participants completed the race version of the IAT. In the IAT, participants used two keys to categorize 12 target images (six Black faces, six White faces) and 16 evaluative words (8 pleasant, 8 unpleasant). They first completed two 20-trial practice blocks, in which they discriminated pleasant from unpleasant words, and Black from White faces. The third and fourth blocks were the critical blocks consisting of 20 and 40 trials, respectively. Participants were instructed to press one key whenever they saw a picture of a White person or a pleasant word, and another key whenever they saw

[2] This study reports the results of a process dissociation procedure (see Payne, 2001) that examined whether Black and White individuals differ in their degree of automatic and controlled processing. This analysis did not, however, test the possibility that Blacks and Whites differ in their level of self-regulation of associations, because the influence of this process cannot be estimated using the process dissociation procedure. Although the controlled component of the process dissociation procedure (C) has, at times, been described as reflecting "self-regulation," it does not measure the ability to regulate automatic associations that have already been activated. Instead, it represents an "early" control process that prevents the activation of automatic bias in the first place.

a picture of a Black person or an unpleasant word. The keys used to categorize Black and Whites faces were switched in the remaining blocks. The fifth block was a practice block in which participants discriminated Black from White faces. In the last two blocks, "Black" shared a response key with the evaluative dimension "pleasant." Participants who respond more quickly when "Black" shares a key with "unpleasant" ("compatible" trials) than when it shares a key with "pleasant" ("incompatible" trials) are thought to have an implicit preference for Whites relative to Blacks (Greenwald et al., 1998).[3]

Target category and attribute labels remained on the top left and top right of the screen throughout the task, while stimulus pictures and words appeared at the center of the screen. The order of the critical blocks was counterbalanced across participants. A red "X" appeared whenever participants made an error, and they were required to correct the error before moving onto the next trial. Latencies were recorded to the correct response. Participants were instructed to make their classifications as quickly and accurately as possible.

Results

IAT Bias

IAT scores were calculated according to the algorithm described by Greenwald, Nosek, and Banaji (2003). Higher IAT effects indicate stronger implicit pro-White preference. Replicating previous findings (e.g., Nosek et al., 2002; Stewart et al., 2009), White participants showed the typical pro-White bias, ($M = 0.40$, $SD = 0.38$) and had significantly higher IAT scores than Black participants ($M = -0.02$, $SD = 0.43$), $t(18,686) = 47.79$, $p < .001$.

Modeling

To explore the processes responsible for the difference between Black and White participants in their IAT performance, we calculated Quad model parameter estimates for each participant. The structure of the Quad model is depicted as a processing tree in Figure 1. In the tree, each path represents a likelihood. Processing parameters with lines leading to them are conditional upon all preceding parameters. For instance, overcoming bias (OB) is conditional upon both association activation (AC) and detection (D).

The conditional relationships described by the model form a system of equations that predict the numbers of correct and incorrect responses in different conditions (e.g., compatible and incompatible trials). For example, a Black target in an incompatible trial will be responded to correctly with the probability: $AC \times D \times OB + (1 - AC) \times D + (1 - AC) \times (1 - D) \times G$. 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 – $(1 - AC) \times D$ – represents the likelihood that the association is not activated, *and* that the correct response can be detected. Finally, $(1 - AC) \times (1 - D) \times G$ represents the likelihood that the association is not activated, *and* that the correct answer cannot be detected, *and* that the participant guesses by pressing the positive ("pleasant") key. Because the "pleasant" and "Black" categories share the same response key in the incompatible block, pressing the positive key in response to a Black target returns the correct answer. 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 to the actual data to determine the model's ability to account for the data. A χ^2 estimate is computed for the difference between the predicted and observed errors. In order 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 χ^2. The final parameter values that result from this process are interpreted as relative levels of the processes.

For each participant, we calculated parameter estimates of association activation, detection, overcoming bias, and guessing. The G parameter was coded so that higher scores represented a bias toward guessing with the positive ("pleasant") key. Two separate AC parameters were estimated: One measuring the extent to which associations between "Black," and "unpleasant" were activated in performing the task, and another measuring the extent to which associations between "White" and "pleasant" were activated. The ability to generate independent estimates of Black–unpleasant and White–pleasant associations is one of the strengths of using the Quad model.[4]

[3] We refer to the Black + unpleasant/White + pleasant pairing as "compatible" because the vast majority of participants respond more quickly and accurately to it than to the Black + pleasant/White + unpleasant pairing.

[4] The implementation of the Quad model used in the current studies is slightly different than the one described by Conrey et al. (2005). The present version of the model estimates a single OB parameter, rather than two OB estimates, as in Conrey et al. (2005). In the current version, the OB parameter is derived only from trials with Black and White targets, and estimates the extent to which associations are overcome in responding to those targets. Theoretically, we believe that participants are much more likely to be motivated and/or able to overcome bias on target trials than on attribute trials because the implications for expressing bias are more obvious on target trials. Consistent with this proposal, the data from both studies show that IAT bias as expressed in greater errors on incompatible than compatible trials (which is the IAT bias modeled by the Quad model) was larger for attribute (Study 1 $M = 0.06$; Study 2 $M = 0.13$) than target (Study 1 $M = 0.01$; Study 2 $M = 0.06$) trials, both p values $< .01$. This is consistent with the suggestion that participants are regulating associations on target trials more so than on attribute trials.

A

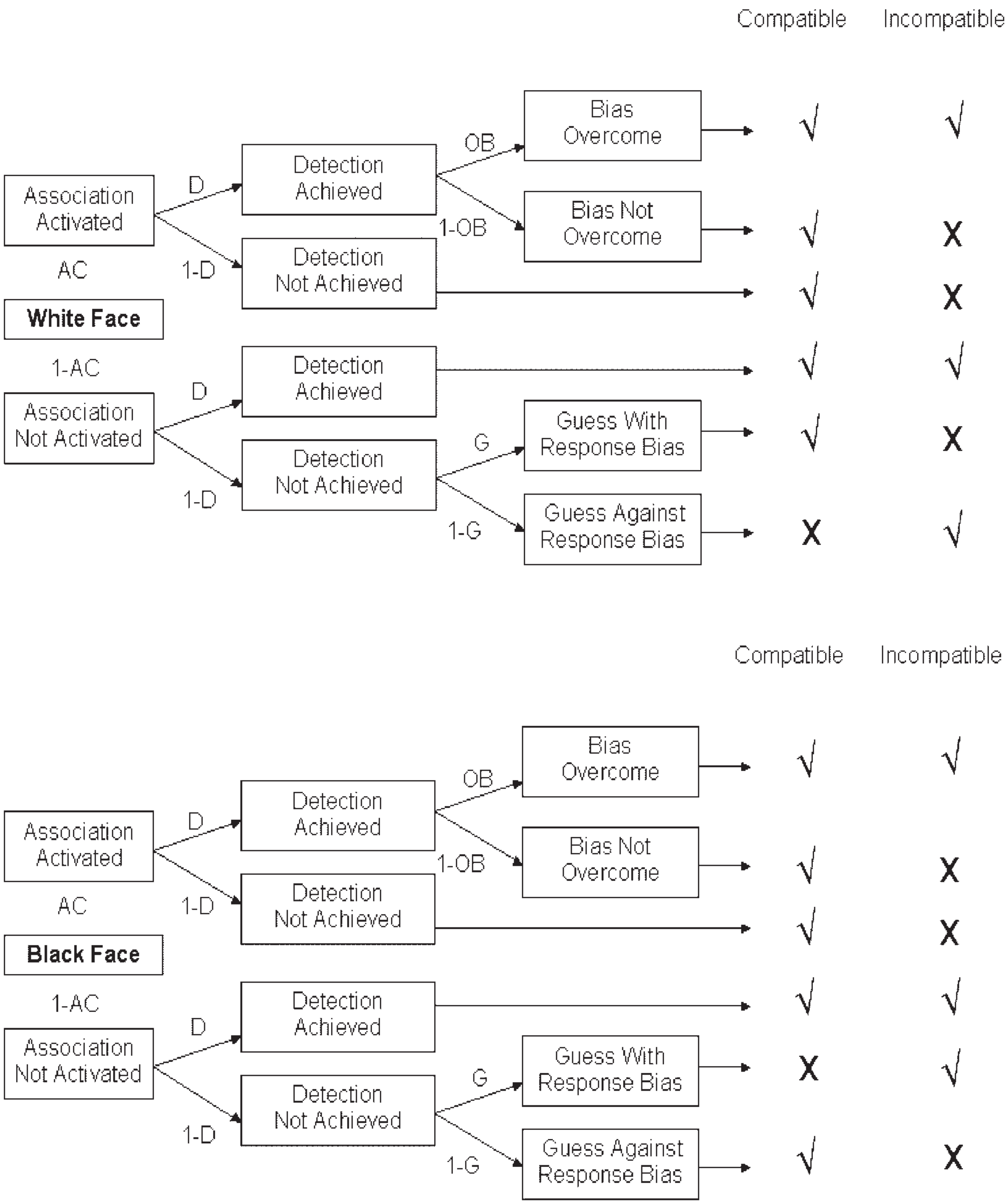

Figure 1. The Quadruple Process model (Quad model). Each path represents a likelihood. Parameters with lines leading to them are conditional upon all preceding parameters. The table on the right side of the figure depicts correct (✓) and incorrect (X) responses as a function of process pattern and trial type (Panel A for targets and Panel B for attributes). In this particular figure, the guessing bias refers to guessing with the positive ("pleasant") key.

B

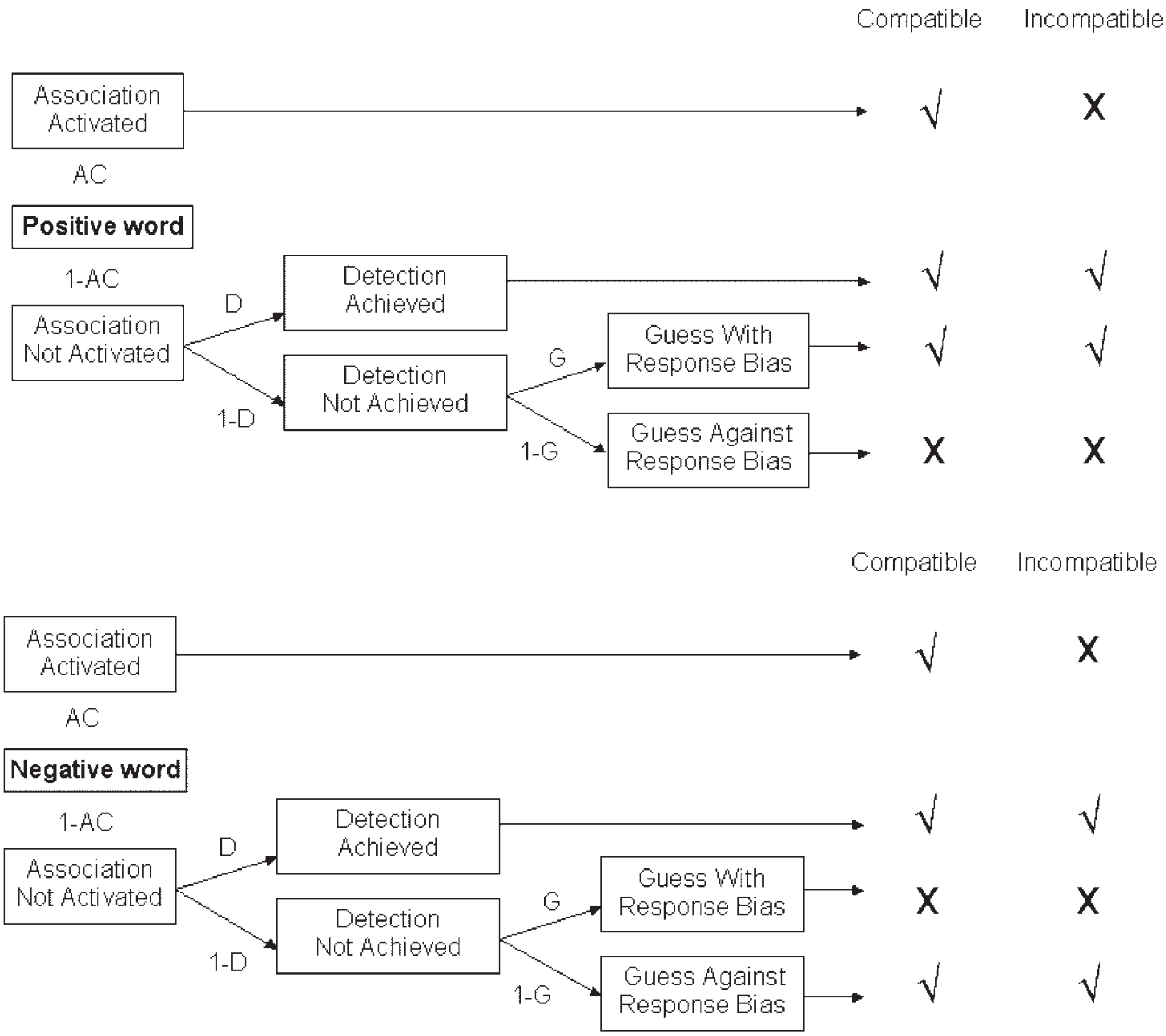
Compatible
Incompatible
Association Activated
√
X
AC
Positive word
1-AC
Association Not Activated
D
1-D
Detection Achieved
√
√
Detection Not Achieved
G
1-G
Guess With Response Bias
√
√
Guess Against Response Bias
X
X
Compatible
Incompatible
Association Activated
√
X
AC
Negative word
1-AC
Association Not Activated
D
1-D
Detection Achieved
√
√
Detection Not Achieved
G
1-G
Guess With Response Bias
X
X
Guess Against Response Bias
√
√

Table 1. Parameter estimates for black-white IAT, Study 1

Parameter		Estimate [*SD*]	
		Black participants	White participants
AC	Black–unpleasant	.04 [.07]	.07 [.09]
	White–pleasant	.05 [.08]	.10 [.11]
D		.88 [.12]	.91 [.10]
OB		.72 [.43]	.73 [.40]
G		.55 [.32]	.56 [.36]

Notes. AC = activation of associations, D = detection, OB = overcoming bias, G = guessing.

The overall error rate for the IAT was 7.7%. The model fit the data for 85.4% of the participants, χ^2 (3) < 7.815.[5] Parameter estimates for the two groups are displayed in Table 1. Analyses showed that there were significant differences in the association activation parameters. Collapsed across both AC parameters, White participants (M = 0.09) had stronger pro-White/anti-Black associations than did Black (M = 0.04) participants, $t(18,686) = 23.776$, $p < .001$, $d = .35$. White participants had higher estimates than Black participants for both the Black–unpleasant associations (M values = 0.07 versus 0.04), $t(18,686) = 15.667$, $p < .001$, $d = .23$ and the White–pleasant associations (M values = 0.10 versus 0.05), $t(18,686) = 22.473$, $p < .001$, $d = .33$. Unexpectedly, White participants also had higher estimates than Black participants for the Detection parameter (Ms = 0.91 versus 0.88), $t(18,686) = 12.137$, $p < .001$, $d = .18$, indicating that they were more able to detect correct and incorrect responses on the IAT. The other parameter estimates did not differ between the groups, p values > .16.

Discussion

Study 1 revealed that White participants had stronger activation of White–pleasant and Black–unpleasant associations than did Black participants. These findings are consistent with Stewart et al.'s (2009) process dissociation analysis, which showed that Black and White participants differed in their degree of automatic, but not controlled, processing. However, use of the Quad model showed that there were group differences in both negative Black and positive White associations. The Quad model also provides an estimate of automatic bias that is uncontaminated by success at overcoming bias (which is measured separately with the OB parameter), and is not dependent on the failure of an "early" controlled component (i.e., D in the Quad model, C in PD models; for a review, see Sherman et al., 2008).

An unexpected finding was that White participants were more able to discriminate between appropriate and inappropriate responses on the IAT. Although speculative, a possible reason for this result is that White people may be adept at monitoring their behavior in race-related situations (Amodio, Devine, & Harmon-Jones, 2008; Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002). This is a question for future research. The important point here is that, despite being more able to detect correct responses than Black participants, White participants showed greater implicit bias on the IAT. Thus, it would seem that any advantage due to enhanced detection (which should produce less bias) was more than offset by the greater activation of biased associations among White participants. There was no evidence that Black and White participants differ in the extent to which they overcome their biased associations.

The finding that Black participants showed less pro-White IAT bias than White participants is consistent with research demonstrating that ingroup contact predicts the extent of implicit intergroup bias (Ashburn-Nardo et al., 2007; Greenwald et al., 1998). One likely consequence of higher ingroup contact is greater exposure to counterstereotypic exemplars. Specifically, because subcultures within a society tend to emphasize positive ingroup exemplars (e.g., Simonton, 1998), Black people may have greater exposure to positive Black exemplars compared with White people. At the same time, Black people may be exposed to more negative White exemplars than are White people, due to encounters with majority group discrimination or via negative depictions of White people in Black culture. Thus, greater exposure to positive Black exemplars and negative White exemplars may have contributed to the Black participants' lower levels of pro-White bias. Furthermore, because estimates of positive White and negative Black associations were lower among Black participants, the current results suggest that any relationship between exemplar exposure and implicit intergroup attitudes may be due, specifically, to activation of automatic associations. However, our conclusions about the effects of exposure to counterstereotypic exemplars on automatic associations are limited because the findings were based on group difference data. We manipulated exemplar exposure in Study 2 to directly test whether exposure to exemplars that are inconsistent with prejudicial associations would reduce the extent to which biased associations are activated.

Study 2

Study 2 examines whether White participants' activation of associations is altered through virtual contact with exemplars. Half of the participants were randomly assigned to a condition in which they were presented with positive Black and negative White exemplars. We predicted that this type of exposure would reduce bias on the IAT relative to a control condition (e.g., Dasgupta & Greenwald, 2001;

[5] The analyses presented in the text include all participants. When participants showing lack of fit were excluded from analysis, the results were unchanged, with one exception. Surprisingly, White participants showed higher levels of OB than Black participants, $t(15963) = 2.39$, $p < .05$, though this effect was weak, $d = .04$.

Govan & Williams, 2004; Mitchell et al., 2003). More importantly, we examined the processes associated with this effect. If the effect is related, specifically, to changes in association activation, participants viewing positive Black and negative White exemplars should exhibit less activation of Black–unpleasant and White–pleasant associations compared with control participants. Alternatively, if seeing valenced exemplars causes the participants to work harder to regulate their biases (see Allen et al., in press; Maddux et al., 2005), overcoming bias should be higher in the experimental condition than in the control condition.

Method

Fifty White undergraduates (39 females) participated in exchange for course credit. Borrowing a paradigm used by Govan and Williams (2004), we asked participants to complete one of two versions of the IAT. Half of the participants were assigned to the control condition. These control participants completed a typical IAT, in which the stimuli were faces of unknown individuals for the Black and White categories. The other participants completed a modified version of the IAT. The face stimuli for this IAT were of five popular Black men (e.g., Martin Luther King, Michael Jordan) and five disliked White men (e.g., Adolph Hitler, Charles Manson). Although earlier pretesting indicated that these individuals were readily recognizable, we wanted to ensure that all participants were aware of the achievements or notoriety of each person, even if they did not recognize his face or name. Thus, in an instruction screen for the IAT, we presented each face together with the individual's name and a brief biographical description (e.g., "Martin Luther King: Leader of the Black Civil Rights Movement in the 1960s"). All other stimuli and procedures were the same as those used in the typical IAT. The block and trial structure of both IATs were identical to the IAT used in Study 1.

Results

IAT Bias

IAT scores were calculated in the same manner as in Study 1. Participants who completed the IAT containing popular Black and disliked White exemplars showed less racial bias ($M = 0.36$, $SD = 0.41$) than the participants who completed the typical IAT ($M = 0.70$, $SD = 0.34$), $t(47) = 3.16$, $p < .01$.

Modeling

The overall error rate was 12.65%. The model fit the data for 90% of the participants, χ^2 (3) < 7.815.[6] Parameter estimates are displayed in Table 2. Exposure to valenced exemplars led to a reduction in the AC parameters ($M = 0.11$) compared to the control condition ($M = 0.17$), $t(47) = 2.08$, $p < .05$, $d = .61$. Follow-up analyses indicated that participants who were exposed to images of popular Blacks and disliked Whites had significantly lower estimates of Black–unpleasant ($M = 0.09$) associations than control participants ($M = 0.15$), $t(47) = 1.97$, $p = .05$, $d = .58$, and marginally lower estimates of White–pleasant ($M = 0.13$) associations than control participants ($M = 0.19$), $t(47) = 1.63$, $p = .10$, $d = .47$. There were no other parameter differences between the groups, p values > .50.

Table 2. Parameter estimates for black-white IAT, Study 2

Parameter		Estimate [*SD*]	
		Positive black and negative white exemplars	Control
AC	Black–unpleasant	.09 [.07]	.15 [.13]
	White–pleasant	.13 [.12]	.19 [.15]
D		.82 [.15]	.85 [.15]
OB		.67 [.47]	.58 [.45]
G		.53 [.32]	.49 [.37]

Notes. AC = activation of associations, D = detection, OB = overcoming bias, G = guessing.

Discussion

In Study 2, we presented White participants with images of popular Black and disliked White targets or of unknown Black and White targets. Participants who were exposed to valenced exemplars showed less IAT bias and activation of Black–unpleasant and White–pleasant associations than participants in the control condition. Importantly, there were no differences between conditions in overcoming bias. These findings do not support the idea that participants work harder to regulate their automatic associations in response to counterprejudicial exemplars, and indicate instead that such exposure decreases the activation of biased associations (Dasgupta & Greenwald, 2001; Dasgupta & Rivera, 2008).

General Discussion

In this paper, we used the Quad model to understand the processes underlying group-based differences in implicit attitudes and malleability of implicit attitudes resulting from exposure to group exemplars. In Study 1, Black participants had lower levels of pro-White IAT bias and less activation of Black–unpleasant and White–pleasant associ-

[6] The analyses presented in the text include all participants. When participants showing lack of fit were excluded from analysis, the results were unchanged. Though the p values for the collapsed AC, the Black AC, and the White AC changed from .04, .05, and .10 to .06, .08, and .13, respectively, the effect sizes showed little to no change: .61, .58, and .47 versus .60, .55, and .47.

ations than did White participants. In Study 2, participants who were exposed to positive Black and negative White exemplars showed less implicit racial bias and less activation of biased associations than participants who were exposed to unknown exemplars. Across two studies, the group that had (or is assumed to have) greater exposure to counterprejudicial exemplars (Black participants in Study 1 and experimental participants in Study 2) exhibited lower levels of implicit racial bias and association activation. There were no differences in overcoming bias in either study, suggesting that self-regulatory efforts were not associated with the effects on implicit bias. The findings suggest that group membership effects and exemplar exposure effects share the same mechanisms: Both result from activation of more or less favorable group associations.

This research illustrates how the Quad model can be used to test competing explanations for variability in and malleability of implicit attitudes. Previous research has found that Black individuals show lower levels of bias on the IAT than White individuals (Nosek et al., 2002). Other research has found that short-term exposure to positive Black exemplars reduces bias on the IAT (e.g., Dasgupta & Greenwald, 2001; Govan & Williams, 2004; Mitchell et al., 2003). Taken on their own, these findings do not reveal whether the effects were due to increased accessibility of counterprejudicial associations (e.g., Dasgupta & Greenwald, 2001), greater self-regulation of prejudiced associations (e.g., Allen et al., in press; Maddux et al., 2005; Richeson et al., 2003), or a combination of these processes. Distinguishing among these accounts is not possible without empirically disentangling the contributions of automatic associations and self-regulation in implicit attitudes. In the absence of the Quad model findings, the data from the implicit measure could be interpreted in multiple ways. Providing a means to tease apart multiple possible interpretations is one of the Quad model's strengths.

These findings extend previous research by revealing that automatic associations underlie short-term and long-term exemplar exposure effects. It is not clear whether exemplar exposure changes the nature of the underlying associations or the particular associations that are temporarily accessible. Though speculative, it is likely that short-term exposure effects are a function of the associations that are activated by positive Black exemplars, such that the accessibility of negative associations with Blacks is temporarily reduced. In contrast, it is likely that differences between Black and White individuals reflect stable differences in the associations that these individuals have with Blacks and Whites. Whether repeated short-term exposure leads to long-lasting change in the nature of Whites' underlying associations will be an important question for future research and for identifying successful prejudice reduction strategies.

The findings also have implications for understanding the mechanisms underlying the effects of contact on intergroup attitudes. Recent meta-analytic evidence indicates that although "optimal" contact conditions (Allport, 1954) facilitate reduction of explicit prejudice through contact, they are not essential for positive outcomes to occur. This finding prompted Pettigrew and Tropp (2006) to propose that mere exposure (Zajonc, 1968) may be responsible for contact effects. That is, contact reduced prejudice because it increases familiarity with an outgroup, which breeds liking for that group. The current findings indicate that positive outgroup contact in the form of brief exposure to positive exemplars reduces activation of biased associations. However, the mere exposure explanation implies that biased automatic associations should decrease as a function of contact alone, leading to the prediction that over time, exposure even to nonvalenced outgroup exemplars should reduce activation of biased associations. This interesting question could be tested in future applications of the Quad model.

Finally, it is worth considering differences between minority groups in their extent of ingroup bias. As noted previously, African Americans do not consistently display implicit ingroup bias. Given that African Americans are both a sizable minority population, but also tend to be socially isolated by de facto segregation in terms of schooling and housing, etc., African Americans could be expected to demonstrate greater cohesion and pro-ingroup bias than other racial minorities who may not have such extensive contact with other members of their own racial ingroup (e.g., Frey & Farley, 1996). We suggest that the degree to which individuals demonstrate a pro-ingroup bias is affected by multiple factors, including the extent to which negative stereotypes are evident in the majority culture. Thus, although African Americans and Japanese Americans are both stigmatized, stigmatization of African Americans may be far greater. It is also important to note that, although exposure to positive Black exemplars reduced the extent of pro-White bias among White participants in Study 2, the bias was not eliminated. This suggests that other factors besides contact may be needed to extinguish racial biases that are ingrained in the dominant society.

Acknowledgments

This research was supported by a grant from the National Science Foundation (BCS 0820855) to the third author. We are grateful to Brian Nosek, Tony Greenwald, and Mahzarin Banaji for sharing data from Project Implicit.

References

Aberson, C. L., Shoemaker, C., & Tomolillo, C. (2004). Implicit bias and contact: The role of interethnic friendships. *Journal of Social Psychology, 144*, 335–347.

Allen, T. J., Sherman, J. W., & Klauer, K. C. (in press). Social context and the self-regulation of implicit bias. *Group Processes and Intergroup Relations*.

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

Amodio, D. M., Devine, P. D., & Harmon-Jones, E. (2008). Individual differences in the regulation of race bias: The role of conflict detection and neural signs for control. *Journal of Personality and Social Psychology*, *94*, 60–74.

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science, 15*, 88–93.

Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. (2003). Black Americans' implicit racial associations and their implications for intergroup judgment. *Social Cognition, 21*, 61–87.

Ashburn-Nardo, L., Monteith, M. J., Arthur, S. A., & Bain, A. (2007). Race and the psychological health of African Americans. *Group Processes & Intergroup Relations, 10*, 471–491.

Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology, 81*, 789–799.

Bartholow, B. D., Dickter, C. L., & Sestir, M. A. (2006). Stereotype activation and control of race bias: Cognitive control of inhibition and its impairment by alcohol. *Journal of Personality and Social Psychology, 90*, 272–287.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6*, 57–86.

Beer, J. S., Stallen, M., Lombardo, M. V., Gonsalkorale, K., Cunningham, W. A., & Sherman, J. W. (2008). The Quadruple model approach to examining the neural underpinnings of prejudice. *NeuroImage*, *43*, 775–783.

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

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

Bodenhausen, G. V., Schwarz, N., Bless, H., & Wanke, M. (1995). Effects of atypical exemplars on racial beliefs: Enlightened racism or generalized appraisals. *Journal of Experimental Social Psychology*, *31*, 48–63.

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.

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.

Dasgupta, N., & Rivera, L. M. (2008). When social context matters: The influence of long-term contact and short-term exposure to admired outgroup members on implicit attitudes and behavioral intentions. *Social Cognition*, *26,* 54–66.

Devine, P. G., Plant, A. E., & Amodio, D. M., & Harmon-Jones, & Vance, S. L. (2002). The regulation of implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology, 82,* 835–848.

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

Frey, W. H., & Farley, R. (1996). Latino, Asian, and Black segregation in U. S. metropolitan areas: Are multiethnic metros different? *Demography*, *33*, 35–50.

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2009). Aging and prejudice: Diminished regulation of automatic race bias among older adults. *Journal of Experimental Social Psychology*, *45*, 410–414.

Gonsalkorale, K., von Hippel, W., Sherman, J. W., & Klauer, K. C. (2009). Bias and regulation of bias in intergroup interactions: Implicit attitudes toward Muslims and interaction quality. *Journal of Experimental Social Psychology*, *45*, 161–166.

Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology, 40*, 357–365.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality & 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.

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology, 53*, 575–604.

Jellison, W. A., McConnell, A. R., & Gabriel, S. (2004). Implicit and explicit measures of sexual orientation attitudes: Ingroup preferences and related behaviors and beliefs among gay and straight men. *Personality and Social Psychology Bulletin*, *30*, 629–642.

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*, 881–920.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*, 774–788.

Lemm, K. M. (2006). Positive associations among interpersonal contact, motivation, and implicit and explicit attitudes toward gay men. *Journal of Homosexuality, 51*, 79–99.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81*, 842–855.

Maddux, W. W., Barden, J., Brewer, M. B., & Petty, R. E. (2005). Saying no to negativity: The effects of context and motivation to control prejudice on automatic evaluative responses. *Journal of Experimental Social Psychology, 41,* 19–35.

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

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6*, 101–115.

Olson, M. A., Crawford, M. T., & Devlin, W. (2009). Evidence for the underestimation of implicit in-group favoritism among low status groups. *Journal of Experimental Social Psychology, 45,* 1111–1116.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*, 653–667.

Payne, B. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81*, 181–192.

Perdue, C. W., Dovidio, J. F., Gurtman, M. B., & Tyler, R. B. (1990). Us and them: Social categorization and the process of intergroup bias. *Journal of Personality and Social Psychology, 59*, 475–486.

Pettigrew, T., & Tropp, L. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*, 751–783.

Richeson, J. A., & Ambady, N. (2003). Effects of situational power on automatic racial prejudice. *Journal of Experimental Social Psychology, 39*, 177–183.

Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S. et al. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nature Neuroscience, 6*, 1323–1328.

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science, 14*, 287–290.

Richeson, J. A., Trawalter, S., & Shelton, J. N. (2005). African Americans' implicit racial attitudes and the depletion of executive function after interracial interactions. *Social Cognition*, *23*, 336–352.

Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. K. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition, 17*, 1–29.

Shelton, J. N. (2003). Interpersonal concerns in social encounters between majority and minority group members. *Group Processes and Intergroup Relations, 6*, 171–185.

Sherman, J. W. (2006). On building a better process model: It's not only how many, but which ones and by which means? *Psychological Inquiry*, *17*, 173–184.

Sherman, J. W. (2009). Controlled influences on implicit measures: Confronting the myth of process-purity and taming the cognitive monster. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new wave of implicit measures* (pp. 391–426). Hillsdale, NJ: Erlbaum.

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.

Simonton, D. K. (1998). Achieved eminence in minority and majority cultures: Convergence versus divergence in the assessments of 294 African Americans. *Journal of Personality and Social Psychology, 74*, 804–817.

Stewart, B. D., von Hippel, W., Radvansky, G. A. (2009). Age, race, and implicit prejudice: Using process dissociation to separate the underlying components. *Psychological Science*, *20*, 164–168.

Tam, T., Hewstone, M., Harwood, J., Voci, A., & Kenworthy, J. (2006). Intergroup contact and grandparent-grandchild communication: The effects of self-disclosure on implicit and explicit biases against older people. *Group Processes & Intergroup Relations, 9*, 413–429.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey CA: Brooks/Cole.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1–27.

Karen Gonsalkorale

School of Psychology
Brennan MacCallum Building (A18)
The University of Sydney
NSW 2006
Australia
Tel. +61 2 9-351-8930
Fax +61 2 9-036-5223
E-mail karen.gonsalkorale@sydney.edu.au

