Journal of Experimental Social Psychology 45 (2009) 161–166

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

Reports

Bias and regulation of bias in intergroup interactions: Implicit attitudes toward Muslims and interaction quality ☆

Karen Gonsalkorale [a,*], William von Hippel [b], Jeffrey W. Sherman [c], Karl Christoph Klauer [d]

[a] School of Psychology, University of Sydney, Brennan MacCallum Building (A18), Sydney, NSW 2006, Australia
[b] School of Psychology, University of Queensland, St Lucia, QLD 4072, Australia
[c] Department of Psychology, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA
[d] Institut für Psychologie, Albert-Ludwigs-Universität Freiburg, D-79085 Freiburg, Germany

ARTICLE INFO

Article history:
Received 30 March 2008
Revised 30 July 2008
Available online 20 August 2008

Keywords:
Intergroup interaction
Implicit attitudes
Automatic associations
Self-regulation

ABSTRACT

Previous research suggests that automatically activated bias manifests itself in behavior that can jeopardize the quality of intergroup interactions. However, regulation of automatic associations has the potential to attenuate their influence on intergroup interaction. To test this possibility, 46 non-Muslim White participants interacted with a Muslim confederate and completed an implicit measure of attitudes toward Muslims. The Quadruple Process model [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] was applied to the implicit measure to estimate participants' strength of negative associations with Muslims and their ability to overcome those negative associations. The confederate's ratings of how much he liked the participants were predicted by an interaction between automatic negative associations and the ability to overcome them. Specifically, when the strength of participants' negative associations with Muslims was low, participants' level of overcoming bias was unrelated to the confederate's ratings. In contrast, the ability to regulate automatic negative associations predicted greater liking when those associations were strong.

The biases that people hold, even those that they may prefer not to have, can lead them to behave negatively toward outgroup members. Evidence for this link has emerged with research employing implicit attitude measures, which seek to capture automatically activated biases that people may not personally endorse or even be aware of having (see Fazio & Olson, 2003). Compared to those who do not show bias on implicit measures, White individuals with high levels of implicit race bias are more likely to display nonverbal discomfort during cross-race interactions (Dovidio, Kawakami, & Gaertner, 2002; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; McConnell & Leibold, 2001), be rated as unfriendly by Black experimenters and confederates (Fazio, Jackson, Dunton, & Williams, 1995; McConnell & Leibold, 2001; Sekaquaptewa, Espinoza, Thompson, Vargas, & von Hippel, 2003), and have interracial roommate relationships that are short-lived (Towles-Schwen & Fazio, 2006). Similar patterns have been found whether bias was assessed using the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998; e.g., McConnell & Leibold, 2001; for a review, see Greenwald, Poehlman, Uhlmann, & Banaji, in press), evaluative priming tasks (e.g., Dovidio et al., 1997), or implicit stereotyping measures (e.g., Sekaquaptewa et al., 2003). These findings indicate that implicit attitudes and beliefs are important predictors of how people behave in face-to-face interactions.

Although the link between implicit biases and intergroup behavior has been established empirically, less is known about how and when this relationship emerges. One possibility is that automatic biases "leak" through behavior in attitude-relevant situations (e.g., Dovidio et al., 2002). According to this account, if an individual associates an outgroup with negativity, this association will precipitate prejudice-consistent nonverbal behavior in the presence of an outgroup member, even when the actor might prefer to act differently. However, when people have sufficient motivation and cognitive resources to regulate the effects of automatically activated associations, they may be able to prevent their biases from influencing behavior (Fazio & Towles-Schwen, 1999). Research indicates that individuals engage in self-control when interacting with or perceiving outgroup members (Richeson & Shelton, 2003; Richeson et al., 2003), who in turn appear to appreciate these self-control efforts. For example, White participants who were instructed to "try not be prejudiced" during an

☆ This research was supported by an Australian Postgraduate Award to the first author and a grant from the Australian Research Council. We thank Brad Hill for running the experimental sessions and the confederates who took part in the study.
* Corresponding author. Fax: +61 2 9036 5223.
E-mail address: kareng@psych.usyd.edu.au (K. Gonsalkorale).

0022-1031/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2008.07.022

upcoming interracial interaction were rated more positively by their Black interaction partners than Whites who were not given these instructions (Shelton, 2003). In another study (Shelton, Richeson, Salvatore, & Trawalter, 2005), Whites who were high in implicit racial bias were evaluated more favorably than their low-bias counterparts because the former were perceived to be more engaged in the interaction. These studies provide indirect evidence that regulation of the expression of prejudice may aid smooth intergroup interactions (e.g., Richeson & Shelton, 2007).

Additional support for the importance of self-regulation in intergroup behavior comes from a recent study in which heterosexual participants interacted with a male confederate whom they believed to be gay (Dasgupta & Rivera, 2006). Results indicated that the relationship between participants' implicit attitudes and their behavior during the interaction was moderated by individual differences in egalitarian motives and self-reported control over behavior in social interactions. Among participants who were low in egalitarian beliefs, those with greater implicit bias exhibited more negative behavior when they were low, but not when they were high, in self-reported behavioral control. When low-egalitarian participants were also low in implicit bias, their behavior toward the confederate did not depend on behavioral control. Thus, implicit prejudice appears to bias behavior only when individuals are both unmotivated and unable to control their behavior in intergroup contexts.

Despite these advances in identifying the circumstances under which implicit attitudes do and do not relate to behavior during intergroup interactions, important unanswered questions remain. First, the proposed mechanism underlying the observed implicit attitude/intergroup behavior link—that activation of automatic associations drives behavior—has not been tested directly. 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 et al., 2004; Bartholow, Dickter, & Sestir, 2006; Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Payne, 2001; Sherman, in press; Sherman et al., 2008). Because implicit attitude measures are not pure reflections of the automatic associations that are hypothesized to drive behavior in intergroup settings, correlations between scores on these measures and behavior in the presence of outgroup members do not necessarily indicate the influence of those associations.

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 (Amodio et al., 2004; Bartholow et al., 2006; Conrey et al., 2005; 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. This means that correlations between scores on these measures and behavior may reflect the role of associations, the ability to overcome those associations, or both processes. This possibility, which has not been considered in previous research, is important, as it may signal an "upstream," early cognitive process that attenuates the influence of automatic associations and facilitates smooth intergroup interactions, independently of the ability to control behavior during the course of an interaction.

One method to separate the strength of automatic associations from the ability to overcome them is the Quadruple Process model (Quad model; Conrey et al., 2005; 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 a single task (e.g., implicit prejudice measures; for reviews of this approach, see Sherman, 2006; Sherman et al., 2008). According to the model, responses on implicit measures of prejudice 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 are automatically activated when responding to a stimulus. The D parameter reflects a relatively controlled process that discriminates between appropriate and inappropriate responses. Sometimes, the activated associations conflict with the detected correct response. For example, on incompatible trials of implicit attitude measures (e.g., trying to associate outgroup faces with positive words in an IAT), automatic associations (e.g., between outgroups and negativity) conflict with detected correct responses. 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. The Quad model and the construct validity of its parameters have been extensively validated in previous research (see Beer et al., in press; Conrey et al., 2005; Sherman et al., 2008).

Because it produces independent measures of association strength and overcoming bias, the Quad model provides a means to test the unique and interactive effects of these processes on intergroup interaction. By isolating the contribution of automatic associations to performance on implicit measures, the Quad model enables a direct test of the idea that automatic associations influence interactions with outgroup members. At the same time, the Quad model permits a test of whether the immediate regulation of automatic associations, as reflected in responses on implicit measures, is sufficient to influence intergroup interactions. To date, no research has directly examined the role of this kind of "upstream" self-regulation in intergroup interactions. Moreover, the Quad model's ability to generate estimates of these self-regulatory processes from performance on an implicit measure circumvents reliance on self-report measures of self-regulation. Thus, application of the Quad model confers important conceptual and empirical gains over previous approaches to examining the relationship between implicit attitudes and intergroup interactions.

The present research

In the current study, we used the Quad model to examine whether the quality of an intergroup interaction is predicted by activation of automatic associations, the ability to overcome automatic associations, or a combination of these processes. Non-Muslim White participants completed an implicit measure of prejudice toward Muslims and interacted with a Muslim confederate. Our measure of interaction quality was the confederate's impressions of participants' likeability. To examine the unique predictive ability of the implicit measure, we also assessed explicit attitudes toward Muslims and controlled for this measure in our analyses. We focused on attitudes and interactions with Muslims because this group is highly salient in domestic and international relations. To date, no experiments involving actual interactions between Muslims and non-Muslims have been reported in the social psychology literature.

Based on previous findings that implicit attitudes predict intergroup behavior (e.g., Greenwald et al., in press), we predicted that participants with higher levels of implicit prejudice would be liked less by the Muslim confederate. More importantly, by applying the Quad model to the implicit prejudice data, we generated estimates of Activation of Associations and Overcoming Bias to examine their capacity to predict interaction quality. If implicit associations influence intergroup interactions (e.g., Dovidio et al., 2002), then AC should be negatively correlated with the confederate's ratings of

likeability. On the other hand, OB should predict greater likeability if overcoming associations contributes to smooth intergroup interactions. Moreover, if regulation of the effects of automatically activated associations prevents them from affecting behavior, then AC should interact with OB, such that OB predicts the confederate's rating when AC is high, but not when AC is low.

Methods

Participants

Forty-six non-Muslim White undergraduates (31 women and 15 men; *M* age = 20.27, *SD* = 3.71) received course credit for participating in the study.

Materials

Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001)

To measure implicit attitudes toward Muslims, participants were presented with a series of trials in which an image from one of four categories (Muslim, White, good, bad) was presented in the middle of a computer screen. The stimuli were 16 Muslim faces (eight male and eight female), 16 non-Muslim White faces (eight male and eight female), 16 positive images (e.g., puppies, flowers) and 16 negative images (e.g., spiders, snakes). The images communicated that the target persons were Muslim via cues such as style of clothing and face covering. Participants were instructed to press the spacebar as quickly as possible if the picture belonged to a target ("go") category, or to do nothing if it did not. Participants first completed three single categorization blocks with a response deadline of 750 ms, followed by three with a deadline of 600 ms. In these 10-trial blocks, participants discriminated good from bad, bad from good and Muslim from non-Muslim White. The order of the blocks was randomized within each response deadline. Distracter stimuli were drawn from the contrasting category (e.g., non-Muslim White faces for Muslim targets). Images were displayed for the duration of the response deadline, after which time feedback appeared in the form of a green "O" or red "X" (indicating correct and incorrect or slow responses, respectively). If participants responded too slowly on a go trial, the computer automatically moved them onto the next trial. Participants next completed two critical blocks with a 600 ms response deadline. The "go" categories were Muslim and bad in the "compatible" block, and Muslim and good in the "incompatible" block. The critical blocks were completed in a randomized order and each consisted of 80 trials. Participants who display greater accuracy in the compatible block compared to the incompatible block are thought to have implicit biases against Muslims. More specifically, the extent that it is easier to categorize together pictures of Muslim faces and negative concepts than it is to categorize together pictures of Muslim faces and positive concepts reflects the strength of association between Muslims and negativity.

Explicit attitudes

The Social Distance Scale (Bogardus, 1933) was used to assess explicit attitudes toward Muslims. Participants completed this scale, along with a battery of questionnaires that are not relevant to the current study, 2–4 weeks prior to taking part in the experiment. Participants were asked to indicate on 5-point scales the degree to which they were willing (1 = *unwilling*, 5 = *willing*) to engage in a variety of different social relationships (e.g., *have as a neighbor*) *with a Muslim person who speaks English*. Scores on the nine items were reversed and averaged to form the scale (α = .94), such that higher scores represent greater social distance.

Ratings of the interaction

The confederates rated on 5-point scales (1 = *not at all*, 5 = *very much so*) how much they enjoyed each interaction, how much they liked their interaction partner, and how comfortable they felt with him or her. These three items were averaged to create separate scores for ratings by the Muslim confederate (α = .80) and by the non-Muslim White confederate (α = .79), with higher scores corresponding to more favorable evaluations of the participants. The participants also rated the confederates using the same three items. However, these data are not reported, as they are not the focus of this paper.

Procedure

Participants were invited by a non-Muslim White male experimenter to take part in the study. Three Muslim males and four non-Muslim White males were employed as confederates. Two participants and two confederates (one Muslim and one non-Muslim White) were scheduled for each session. The Muslim confederates were of Middle Eastern or South Asian descent. According to their personal preferences, each Muslim confederate was bearded and one wore traditional Muslim dress. After giving informed consent, the participants and confederates were asked to wear a name badge. The experimenter then read each name aloud and checked it off a list. To increase the likelihood that participants would be aware of the confederates' group memberships, the experimenter mispronounced the name (Jalal, Hasan, or Usman) of the Muslim confederate, who corrected him and casually mentioned that it was a common Muslim name.

Participants were then paired with a confederate and led into adjacent rooms containing two chairs. When they were both seated, the experimenter handed each person a list of two topics (e.g., "sport and society" and "gambling"), asked them to pick one of the topics to discuss, and left the room. The confederates were instructed to interact with the participant as they would with anyone else, with the restriction that they not initiate the conversation. Three minutes after the interaction started, the experimenter returned and collected the topic list, ostensibly in preparation for the next part of the experiment. The true purpose of this procedure was to encourage the participant and confederate to engage in a less formal discussion. After a further two minutes, the experimenter stopped the interactions and placed each person in a separate room to complete their ratings of the interaction. Once they had rated the interaction with the Muslim (or non-Muslim) confederate, participants were led into the second room, provided with a second topic list, and asked to interact with the other confederate.[1]

Following the interactions, participants were seated at a computer in a private cubicle. There they completed the GNAT, several tasks that were unrelated to the current study, and demographic questions. Finally, participants were fully debriefed.

Results

Due to procedural issues (e.g., a participant arrived late to the experiment and was not able to interact with both confederates), there were missing data for some measures. This is reflected in the varying degrees of freedom across measures.

Preliminary analyses

GNAT scores were calculated according to the algorithm described by Nosek and Banaji (2001). First, the numbers of hits

[1] Although we counterbalanced the order in which participants interacted with the Muslim and non-Muslim White confederates, we are unable to examine the effects of this variable, as we did not record order information for each participant.

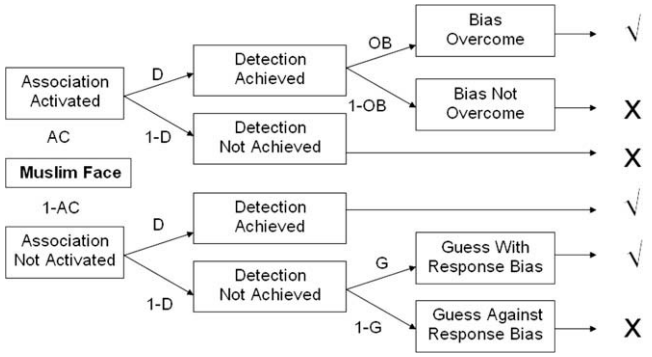

Fig. 1. The Quadruple Process model (Quad model). Each path represents a likelihood. Parameters with lines leading to them are conditional upon all preceding parameters. The column on the right side of the figure depicts correct (√) and incorrect (X) responses as a function of process pattern in the incompatible block of the Muslim GNAT. In this particular figure, the response bias refers to guessing with the spacebar.

and false alarms for each critical block were summed and a correction[2] applied to cells with perfect hits and zero false alarms. Next, the proportions of hits and false alarms for each block were used to calculate d' scores, which reflect sensitivity (ability to discriminate target from distracters). Three participants with negative d' scores for the incompatible block were removed from analysis because these sensitivity values indicate chance responding (Nosek & Banaji, 2001).[3]

A paired-samples t-test indicated that sensitivity was higher in the compatible block (d' = 2.46, SD = .58) than in the incompatible block (d' = 1.41, SD = .52), $t(42) = 9.11$, $p < .0001$. Thus, consistent with previous studies of implicit attitudes toward Muslims (e.g., Park, Felix, & Lee, 2007; Rowatt, Franklin, & Cotton, 2005), participants, on average, showed implicit prejudice toward Muslims.

To examine relationships between implicit attitudes and evaluations of the participants, GNAT difference scores were calculated by subtracting d' values for the incompatible block from the d' values for the compatible block, so that higher scores represent more negative implicit attitudes. Controlling for the explicit attitude measure and whether or not English was the participant's first language[4], GNAT difference scores significantly predicted the Muslim confederate's ratings, $b = -.25$, $\beta = -.30$, $t(39) = -2.10$, $p < .05$, but not the non-Muslim White confederate's ratings, $b = -.03$, $\beta = -.04$, $t(37) = -0.21$, $p = .83$. Thus, participants who displayed greater accuracy in pairing Muslim with bad (relative to good) on the GNAT received less positive evaluations from the Muslim confederate. As GNAT scores did not predict the non-Muslim White confederate's ratings, the remaining analyses focused on elucidating the processes responsible for the relationship between implicit attitudes and the Muslim confederate's ratings.

Quad model analyses: effects of Association Activation and Overcoming Bias

The structure of the Quad model is depicted as a processing tree in Fig. 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 number of correct and incorrect responses in compatible and incompatible trials. The model's predictions are then compared to error rates in the actual data to determine the model's ability to account for the data. A $\chi 2$-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 $\chi 2$. The final parameter values that result from this process are interpreted as relative levels of the four processes. A complete description of data analysis within the Quad model can be found in Conrey et al. (2005). In the current study, frequencies of hits, misses, false alarms, and correct rejections in the GNAT were used to calculate parameter estimates of Association Activation, Detection, Overcoming Bias, and Guessing for each participant. Two separate AC parameters were estimated: one measuring the extent to which associations between "Muslim" and "bad" were activated in performing the GNAT, and one measuring the extent to which associations between "non-Muslim" and "good" were activated. The ability to generate independent estimates of Muslim-bad and non-Muslim-good associations is one of the strengths of using the Quad model. Two separate D parameters were estimated also: one for Muslim faces, and one for non-Muslim faces and good and bad images.[5]

Table 1
Parameter estimates for the Muslim GNAT

Parameter		Estimate [confidence intervals]
AC	Muslim-bad	.14 [.11, .17]
	non-Muslim-good	.15 [.11, .18]
D	Muslim	.84 [.79, .88]
D	non-Muslim, good, bad	.64 [.61, .67]
OB		.32 [.05, .59]
G		.49 [.44, .53]

Notes. Goodness of model fit: $\chi^2(2) = 1.01$, $p = .60$. AC, Activation of Associations; D, Detection; OB, Overcoming bias; G, Guessing.

The Quad model fit the data well, $\chi^2(2) = 1.01$, $p = .60$. Means and confidence intervals for the parameter estimates are displayed in Table 1. Subsequent analyses focused on the roles of the Muslim-bad AC parameter and the OB parameter in predicting the Muslim confederates' ratings. To test the unique and interactive effects of these parameters, we regressed Muslim-bad AC estimates (centered), OB estimates (centered), explicit attitudes (centered), English as a first language (coded 1 for English was the participant's first language and −1 for English was not the first language), and the interaction between AC and OB, on the confederate's ratings. AC, OB, and explicit attitudes were entered in the first step and the interaction term in the second step. Using a web utility (Preacher, Curran, & Bauer, 2006), interactions were followed up

[2] As each block consisted of 80 trials, perfect scores were denoted by zero false alarms or 40 hits for a particular block. Similar to Nosek and Banaji (2001), blocks with zero false alarms were replaced with the value 0.35 and blocks with 40 hits were replaced with the value 39.65.

[3] The results reported in this paper were unchanged when these three participants were included in analysis.

[4] We controlled for English as a first language as it may influence impressions formed during the interaction. The results were the same when we excluded the six participants for whom English was not their native language.

[5] Previous studies (e.g., Conrey et al., 2005) implemented a version of the Quad model that estimates a single D parameter across all item types. This single D version of the model did not provide good fit at the aggregate level for the current GNAT data, $\chi 2(3) = 33.61$, $p < .0001$. In contrast, the two D version fit the data well. This difference in fit likely reflects the special meaning that the target category holds in single-category variants of the IAT. In contrast to the items from the other three categories (non-Muslim, good, bad), items from the target category (Muslim) were always designated as "go" items in the GNAT. Thus, there are likely to be differences in discriminability for the Muslim items compared to the other three item types. Indeed, setting the two D estimates equal to each other resulted in loss of fit, $\Delta\chi 2(1) = 32.60$, $p < .0001$, indicating that the D estimate for Muslim faces was significantly higher than the D estimate for the attributes and non-Muslim faces.

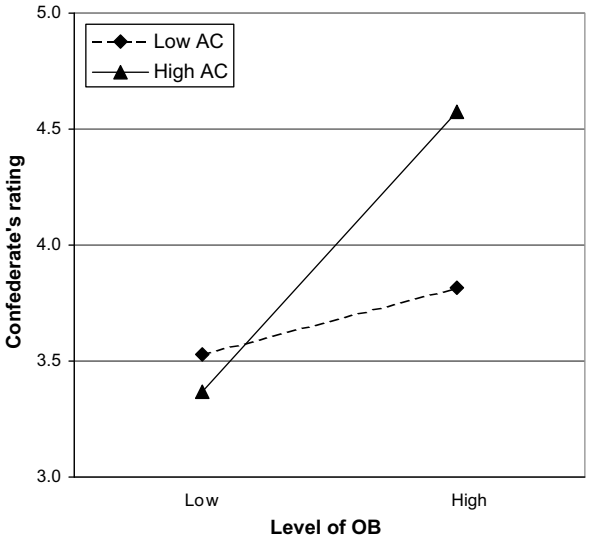

Fig. 2. Regression slopes predicting the Muslim confederate's rating from Overcoming Bias (OB) at one standard deviation above and below the mean of Muslim-bad Activation of Associations (AC). Higher ratings reflect more positive evaluations.

by calculating simple slopes at one standard deviation above and one standard deviation below the mean of AC.

The regression analysis revealed no unique effects of AC, $b = -.65$, $\beta = -.13$, $t(38) = -0.81$, $p = .42$, or OB, $b = .22$, $\beta = .16$, $t(38) = 0.97$, $p = .34$. The explicit attitude measure, $b = -.16$, $\beta = -.22$, $t(38) = -1.45$, $p = .15$, and the English as a first language variable, $b = .23$, $\beta = .26$, $t(38) = 1.68$, $p = .10$, also did not significantly predict likeability ratings. However, as shown in Fig. 2, OB interacted with AC, $b = 4.13$, $\beta = .38$, $t(37) = 2.09$, $p < .05$.[6,7] Tests of simple slopes revealed that OB was unrelated to the confederate's ratings at one standard deviation below the mean of AC, $b = .33$, $t(37) = 1.34$, $p = .19$. In contrast, OB significantly predicted greater likeability at one standard deviation above the mean of AC, $b = 1.37$, $t(37) = 2.11$, $p < .05$. Surprisingly, among participants with high OB, those who had high levels of AC were liked more than those with low levels of AC (see Fig. 2).

Discussion

Some researchers have theorized that automatic associations influence behavior, resulting in discrimination toward outgroup members (Dovidio et al., 2002), even when the actor would prefer to act differently. More egalitarian behavior results from less biased associations. Others have suggested that positive intergroup interactions flow from people's efforts to control their behavior during interactions (Shelton et al., 2005). Still others have proposed that automatically activated associations and behavioral regulation interact to influence behavior in intergroup settings (Dasgupta & Rivera, 2006; Fazio & Towles-Schwen, 1999). In the current study, we explored an alternative source of controlling intergroup interaction, the ability to regulate the effects of automatic associations far "upstream," even before they can influence responses on an implicit measure. We examined the capacity of Quad model estimates of activation of anti-Muslim associations (AC) and the ability to overcome those associations (OB) to predict the quality of an intergroup interaction. Levels of AC and OB when performing the GNAT were not significantly associated with the Muslim confederate's ratings. Instead, the confederate's ratings of how much he liked the participants were predicted by an interaction between AC and OB. Participants' ability to overcome automatic anti-Muslim associations predicted greater likeability only when those associations were strong. When AC was low, OB did not predict likeability ratings. Thus, the quality of the intergroup interaction in this study was best predicted by a combination of AC and OB.

The current study illustrates how the Quad model may enhance interpretability of data from implicit attitude measures. We found that participants who exhibited greater prejudice toward Muslims on the GNAT received less positive ratings from the Muslim confederate. Taken on its own, this result might be interpreted in a variety of ways. For example, approaches that treat implicit attitude measures as pure reflections of automatic associations would conclude that stronger associations predict disliking. The negative relationship between GNAT bias scores and likeability might also be used to refute the importance of self-regulation, as those who are presumed to regulate the most (i.e., those with higher implicit measure bias scores; Shelton et al., 2005) were liked the least. In contrast, our Quad model findings indicate that biased associations alone do not jeopardize the quality of an intergroup interaction. The modeling further demonstrates that regulation of associations plays an important role when people have strong automatic associations. In the absence of the Quad model findings, the data from the implicit measure would lead to very different conclusions. Providing a means to tease apart multiple possible interpretations of effects involving implicit attitudes is one of the Quad model's strengths.

The current findings also extend the literature on the role of self-regulation in intergroup interaction. Previous research (Dasgupta & Rivera, 2006) found that when people are not motivated to be egalitarian and are high in implicit bias, self-reported ability to control behavior within social interactions predicts more positive behavior in the presence of an outgroup member. Our findings show that early, "upstream" self-regulatory processes that directly oppose activated associations may impact the quality of intergroup interactions in a similar way to self-reported regulation of behavior during an interaction. It should be noted that, although our findings indicate that rapid regulation of the effects of automatic associations (as reflected on responses on implicit measures) is sufficient to influence interaction quality, they do not speak to other kinds of self-regulation that may occur during the interaction. Future research could incorporate measures of verbal and nonverbal behavior to explore the relationship between early regulation processes of the sort examined here and within-interaction self-regulation, and how the two contribute to interaction quality. Because we focused on the confederate's perceptions, our results also do not shed light on the behavioral mechanism underlying the effects of AC and OB on interaction quality. In all likelihood, AC and OB probably influence a combination of verbal and nonverbal responses to outgroup members, but this remains a question for future research.

An unexpected finding was that, among participants with high OB, those with high AC were better liked than those with low AC. One possible account of this finding is that the effort required to overcome associations increases with the strength of those associations and, therefore, may be more apparent to interaction partners. The perception that a partner is working hard to control

[6] Individual parameter estimates that were generated using the single D version of the Quad model produced the identical interaction.

[7] As indicated previously, one of the three Muslim confederates chose to wear traditional Muslim dress. An examination of whether confederates wore traditional versus non-traditional attire indicated no unique effects of attire, $p > .20$. The 3-way interaction between AC, OB, and attire also was not significant, $p > .36$. However, the key interaction between AC and OB remained significant in this analysis, $b = 4.23$, $\beta = .39$, $t(36) = 2.17$, $p < .05$. Thus, differences in attire between the Muslim confederates did not appear to influence the results.

bias may be appreciated (e.g., Shelton et al., 2005). It is important to note, however, that highly biased associations, in and of themselves, do not yield partner approval. It is only when highly biased associations are accompanied by strong success in overcoming those associations that partners are held in high regard. Further research will be needed to explore this effect.

A possible limitation of the current study concerns the temporal order of the tasks. Because participants completed the GNAT after the interaction, implicit attitude performance may have been influenced by the encounter with the Muslim confederate. However, studies in which implicit attitudes were measured either immediately prior to the interaction (Dovidio et al., 1997; Fazio et al., 1995; Sekaquaptewa et al., 2003) or in a separate experimental session (Dasgupta & Rivera, 2006; Towles-Schwen & Fazio, 2006) also found that implicit bias correlates with less favorable impressions. In addition, meta-analytic evidence suggests that task order does not moderate the correlation between implicit and explicit measures (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). Moreover, a recent meta-analysis (Greenwald et al., in press) found that task order had no effect on the ability of the IAT to predict behavior. This consistency across studies suggests that the current findings are not dependent on the task order employed.

In summary, both automatic associations and the ability to overcome those associations have been identified as important contributors to intergroup interaction. The Quad model provides a unique tool for separating these processes and ascertaining their relationships with behavior without relying on implicit measures as pure proxies for association strength and without relying on self-report measures of self-regulation. The model also permits tests of a specific, upstream form of self-regulation that is difficult to otherwise measure. Altogether, application of the model provides a more nuanced understanding of the relationship between implicit attitudes and intergroup interaction.

References

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.

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 and Review, 6*, 57–86.

Beer, J. S., Stallen, M., Lombardo, M. V., Gonsalkorale, K., Cunningham, W. A., & Sherman, J. W. (in press). The Quadruple Model approach to examining the neural underpinnings of prejudice. *NeuroImage*.

Bogardus, E. S. (1933). A social distance scale. *Sociology & Social Research, 17*, 265–271.

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., & Rivera, L. M. (2006). From automatic antigay prejudice to behavior: The moderating role of conscious beliefs about gender and behavioral control. *Journal of Personality and Social Psychology, 91*, 268–280.

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

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33*, 510–540.

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

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology, 54*, 297–327.

Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 97–116). New York: Guilford.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (in press). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*.

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.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*, 1369–1385.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37*, 435–442.

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-Go Association Task. *Social Cognition, 19*, 625–666.

Park, J., Felix, K., & Lee, G. (2007). Implicit attitudes towards Arab-Muslims and the moderating effects of social information. *Basic and Applied Social Psychology, 29*, 35–45.

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.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*, 437–448.

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., & Shelton, J. N. (2007). Negotiating interracial interactions: Costs, consequences, and possibilities. *Current Directions in Psychological Science, 16*, 316–320.

Rowatt, W. C., Franklin, L. M., & Cotton, M. (2005). Patterns and personality correlates of implicit and explicit attitudes toward Christians and Muslims. *Journal for the Scientific Study of Religion, 44*, 29–43.

Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., & von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology, 39*, 75–82.

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

Shelton, J. N., Richeson, J. A., Salvatore, J., & Trawalter, S. (2005). Ironic effects of racial bias during interracial interactions. *Psychological Science, 16*, 397–402.

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. (in press). 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*. 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.

Towles-Schwen, T., & Fazio, R. H. (2006). Automatically activated racial attitudes as predictors of the success of interracial roommate relationships. *Journal of Experimental Social Psychology, 42*, 698–705.