Title

Social context and the self-regulation of implicit bias

Journal

Group Processes & Intergroup Relations, 13(2)

ISSN

1368-4302

Authors

Allen, Thomas J
Sherman, Jeffrey W
Klauer, Karl Christoph

Publication Date

2010-03-01

DOI

10.1177/1368430209353635

Peer reviewed

Group Processes & Intergroup Relations

http://gpi.sagepub.com

Social context and the self-regulation of implicit bias

Thomas J. Allen, Jeffrey W. Sherman and Karl Christoph Klauer

Group Processes Intergroup Relations 2010; 13; 137
DOI: 10.1177/1368430209353635

The online version of this article can be found at:
http://gpi.sagepub.com/cgi/content/abstract/13/2/137

Published by:
SAGE
http://www.sagepublications.com

Article

Group Processes & Intergroup Relations
G P I R

Social context and the self-regulation of implicit bias

Group Processes & Intergroup Relations
13(2) 137–149

DOI: 10.1177/1368430209353635
gpir.sagepub.com
SAGE

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

Abstract

Placing outgroup members in positive social contexts can reduce implicit bias. Different contexts may activate different associations of the group. Alternatively, contexts may act as cues for inhibiting bias. We applied the Quad model (Sherman et al., 2008) to address these possibilities. We also examined how motivation moderates these effects. Participants completed a Black–White evaluative priming task with primes presented in positive versus negative contexts and a measure of motivation to control prejudice. Results showed less implicit bias in positive versus negative contexts and that this effect was stronger among highly motivated participants. Modeling revealed that these effects were related to inhibition of biased associations, but not to changing the associations that were activated. Implications for prejudice reduction are discussed.

Keywords

implicit attitudes, inhibition, motivation, prejudice, social context

Paper received 14 February 2009; revised version accepted 21 September 2009.

Since the inception of the civil rights movements of the mid-20th century, social norms regarding the acceptance of racial prejudice have moved in an increasingly egalitarian direction (Rokeach & Ball-Rokeach, 1989). Correspondingly, self-reports of racial prejudice have declined substantially in the last 40 years (Schuman, Steeh, & Bobo, 1985). Despite these positive developments, racial inequality continues to be pervasive in areas of employment, housing, and health care (Pager & Shepherd, 2008; Smedley, Stith, & Nelson, 2003). The discrepancy between improved attitudes and continuing negative outcomes for racial minorities has prompted social psychologists to develop more subtle ways of measuring racial bias (Fazio, Jackson, Dunton, & Williams, 1995; Greenwald, McGhee, & Schwarz, 1998). These new implicit measures have revealed pervasive racial bias, even among those who report little to no racial prejudice on traditional attitude measures (Fazio et al., 1995; Greenwald et al., 1998). Moreover, implicit bias has been shown to predict a variety of discriminatory behaviors, in

[1] University of California, Davis
[2] University of Freiburg

Corresponding author:
Thomas J. Allen or Jeffrey W. Sherman, Department of Psychology, University of California, One Shields Avenue, Davis, CA 95616, USA
[email: tjallen@ucdavis.edu or jsherman@ucdavis.edu]

many cases more effectively than self-report measures (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). In response to these findings, social psychologists have turned to investigating factors that might attenuate implicit bias (e.g., Blair, 2002; Sherman et al., 2008).

Context and implicit bias: Activation of different associations or inhibition?

One factor that researchers have increasingly focused on is the social context in which outgroup targets are encountered. For example, Wittenbrink, Judd, and Park (2001) found that bias on a Black–White Implicit Association Test (IAT; Greenwald et al., 1998) was reduced after viewing African-Americans in a positively valenced context (e.g., family barbeque) as opposed to a negatively valenced context (e.g., gang-related shooting). Similarly, Barden, Maddux, Petty, and Brewer (2004) found that, when African-Americans were depicted in positively valenced settings, such as a factory or church, implicit bias was reduced compared to negatively valenced settings, such as a prison.

Typically, such effects have been explained as being due to the activation of different associations in the different contexts. For example, Wittenbrink et al. (2001) suggested that when a Black male is encountered in a positive setting (e.g., church; family barbeque), positive aspects of the African-American stereotype can be activated, leading to reduced implicit bias. Alternatively, in such contexts, more positive social roles, such as church-goer, might be activated, thereby reducing implicit bias (e.g., Barden et al., 2004).

Other studies have pointed to the role of inhibition processes in producing context-based variations in implicit bias. Based on Monteith's work on prejudice reduction (e.g., Monteith, Ashburn-Nardo, Voils, & Czopp, 2002), Maddux, Barden, Brewer, and Petty (2005) proposed that social contexts can act as cues that a prejudiced response is imminent and prompt inhibition of biased associations. Supporting this view, Maddux et al. (2005) found that participants who were highly motivated to control prejudiced responses were less likely to show implicit race bias toward Black targets presented in negative contexts (e.g., prison) than in positive contexts (e.g., church). Presumably, because motivated participants are more concerned with reducing biased responding, they are more attentive to contextual cues that may signal biased responses than are unmotivated participants.

Determining the extent to which such contextual effects are due to changes in activated associations versus inhibition processes is difficult because responses on implicit measures reflect a combination of activation and control processes. The independent influences of these processes cannot be determined based on task performance alone. Consider the Stroop task. A young child who knows colors but does not know how to read will likely perform very well on the task, making few errors. An adult with full reading ability may achieve the same level of success. However, these performances would be based on very different underlying processes. In the case of the adult, in order to perform the task accurately, the automatic habit to read the word must be overcome in order to report the color of the word accurately. In contrast, the child has no automatic habit to overcome on incompatible trials (e.g., the word "Blue" written in red ink). The same logic applies to implicit measures of attitudes. That is, the same level of implicit bias demonstrated by two people or across two contexts may reflect moderately biased associations in one case, but strong associations that are successfully overcome in the other (e.g., Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Gonsalkorale, Sherman, & Klauer, 2009; Gonsalkorale, von Hippel, Sherman, & Klauer, 2009; Sherman et al., 2008). In the same way, different levels of implicit bias across people or contexts may reflect differences in activated associations, differences in inhibition of those associations, or a combination of both. Thus, although it is possible that the results reported by Wittenbrink et al. (2001) and Barden et al. (2004) reflected the fact that different contexts activated

different sets of associations, it also is possible that the results reflected the fact that different contexts prompted differing levels of inhibition. Similarly, though it is possible that Maddux et al.'s (2005) results reflected the role of context in provoking inhibition, they also may have reflected context-specific variations in associations activated among different participants.

In an attempt to address this issue, we have developed and validated a method to assess the independent contributions of automatically activated associations and inhibition of those associations to implicit task performance (for a review, see Sherman et al., 2008). We used this method, known as the Quadruple Process or "Quad" model, in the present research to estimate, for the first time, the independent contributions of biased associations and inhibition of those associations to contextual variation in implicit race bias.

The Quad model

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, see Sherman, 2006; Sherman et al., 2008). A central feature of the model is that it goes beyond the basic division between automatic processes and controlled processes. In many dual-process theories, the controlled process is one that seeks out, or extracts, information from the environment to provide accurate and appropriate judgments and behaviors (Epstein, 1994; Ferreira, Garcia-Marques, Sherman, & Sherman, 2006; Kahnenman, 2003; Sloman, 1996). In other dual-process theories, the controlled process is self-regulatory in nature. Hence, control is exerted to inhibit or suppress inappropriate or unwanted responses (Baumeister & Vohs, 2004; Carver & Scheier, 1981). Contemporary dual-process models tend to focus on one of these two processes. Although both are reliant on cognitive resources, research has shown that these processes are independent of each other (Miller & Cohen, 2001). It also clear that in many types of behaviors, both processes operate simultaneously. For example, to deliver the correct response on a Stroop task (for a review, see MacLeod, 1991), people must determine both the color of the ink (accuracy), and they must suppress the habit of reading the word (inhibition).

Automatic processes have also been characterized in two different ways. Most commonly, automatic processes are depicted as simple associations that are triggered by environmental stimuli without the perceiver's awareness or intent (e.g., Schneider & Shiffrin, 1977). Such automatic processes may require later correction by controlled processes. In the Stroop task, for example, the automatic habit of reading the word can interfere with the controlled response of naming its color, and the habit may need to be corrected. In other dual-process models, however, automatic processes play a different role. Rather than interfering with controlled processing, automatic processes may act as a secondary source of responding when control fails. For example, prominent dual-process models of memory posit that, when controlled attempts at recollection fail, people may instead rely on automatically generated feelings of familiarity to identify a stimulus as old (Jacoby, 1991).

The four processes described above show up repeatedly in dual-process models of automaticity and control. These models typically include one automatic and one controlled process. However, a model in which all four of these processes are considered and measured can provide a more detailed analysis of the determinants of behavior. Toward this end, we have developed the Quad model. 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 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 detects the nature of the

stimulus and determines the correct task response (i.e., which button to push). Sometimes, the activated associations conflict with the detected correct response. For example, on incompatible trials of implicit attitude measures (e.g., trying to respond to positive terms following a Black prime), 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. Thus, the Quad model synthesizes all four processes that have previously been discussed in dual-process models. One of the most important features of the Quad model is, however, that it goes beyond a mere narrative description of the proposed processes. The basic assumptions of the Quad model have been implemented in a multinomial model that can provide independent quantitative estimates of each of the four processes. The mathematics underlying the model are described in further detail in the results section below. The Quad model and the construct validity of its parameters have been extensively validated in previous research (see Beer et al., 2008; Conrey et al., 2005; Gonsalkorale, Sherman, et al., 2009; Gonsalkorale, von Hippel, Sherman, & Klamer, 2009; Sherman et al., 2008).

Present research

In the current study, participants completed a standard Black–White evaluative priming task (Fazio et al., 1995), with one notable alteration: Black and White primes were presented in positive and negative contexts. Based on previous findings (Barden et al., 2004; Wittenbrink et al., 2001), we predicted that participants would show greater racial bias on the task when the targets were paired with negative contexts than when paired with positive contexts. We also measured motivation to control prejudiced responses (Dunton & Fazio, 1997). Our expectation was that, as motivation increased, participants would be more sensitive to using the contextual cues to reduce biased responding. Thus, we predicted that motivation would accentuate the contextual moderation effect.

We used the Quad model to distinguish between the association activation and association inhibition accounts of this effect. If the effect is due to the influence of the contexts on the associations that are activated, then Quad model estimates of activation of negative associations (AC) of African-Americans should be higher in negative contexts (e.g., prisons) than in positive contexts (e.g., church). Obversely, if association inhibition is responsible for the effect, then Quad model estimates of overcoming anti-Black bias (OB) should be higher in the positive than the negative contexts. Of course, it might be that both activation and inhibition processes account for the effect, in which case, lower Black AC and higher OB may both be observed in positive contexts. Given the demonstrated role of motivation in moderating the effects of context on implicit bias, whichever process is associated with reduced bias also should be associated with higher levels of motivation to respond without prejudice.

Method

Participants Forty-nine non-Black undergraduates at a university in northern California participated in this study for partial course credit.

Materials Twelve color photographs of college-aged White and Black males with neutral expressions were used. All photos were head shots taken against a monochrome background. Photo editing software was used to remove the background and replace it with the experimental contexts. These contexts consisted of interior shots of a church and a prison cell. The original 12 photos were transformed into two sets of images that depicted the identical target standing in each of the two contexts. The contexts were

selected so that a face could plausibly appear in the bottom center of the photo. Three pleasant words (e.g., wonderful, pleasure, celebrate) and three unpleasant words (e.g., agony, horrible, disaster) also were used for the priming task.

Procedure With a few modifications, the procedure followed the standard evaluative priming procedure designed to measure implicit bias (Fazio et al, 1995). Participants were instructed that the first task of interest was to see how fast they could categorize words as pleasant or unpleasant. This first phase of the experiment was designed to obtain participants' baseline response rates to each of the words that would be used as targets in the subsequent priming procedure. For this task, the words, either positive or negative in meaning, were presented and participants were told to press the "E" key for unpleasant words and the "I" key for pleasant words. Six practice trials, consisting of filler words, were presented in order to familiarize participants with the task. The beginning of each trial was signaled by a presentation of a row of asterisks in the center of the computer screen for 500 ms. The asterisks were replaced by a blank screen for 100 ms, after which the word appeared on the screen and remained until the participant responded. An inter-trial delay of 3 s followed each response. After practice trials, the six target words were randomly presented twice each in the same manner. Latencies were recorded for these trials to the nearest millisecond to provide baseline response times for each participant.

Participants were told that the second phase of the experiment was a face recognition task. This task was provided to maintain the cover story for the later priming phase of the experiment (i.e., that the word categorization and memory tasks were ostensibly being combined). Participants were instructed to pay attention to several faces that were to appear briefly on the screen, as they would later be tested for memory of these faces. Eight White and Black male faces were then presented one by one on the screen for 215 ms, followed by an inter-trial delay of 3 s. None of these faces was used in the key priming task that followed. Sixteen trials (two of each face) were presented in random order. After viewing all of the faces, participants were then presented with all of the previous eight faces along with eight new foils. Their objective was to make a decision as to whether the face was one of the previous eight that they were presented with earlier or if the face was new. They made their decision by pressing the "E" key for new faces and the "I" key for old faces.

Phase three of the experiment involved the actual priming task. As part of the cover story, the instructions indicated that we were now interested in how well participants could perform the face recognition and word judgment tasks simultaneously. At the beginning of each trial, a row of asterisks appeared for 500 ms. The asterisks were followed by a blank screen for 100 ms, after which either a Black or White face was presented for 215 ms or 445 ms.[1] After a 100 ms delay, a target word appeared on the screen. Participants were instructed to remember each face for a future recognition test, while at the same time judging the meaning of each target word. Six practice trials were presented, consisting of filler faces and words. The practice trials were then followed by a series of 144 test trials. Filler and test faces were presented in either a church or a jail context. Faces and contexts were presented simultaneously. Pairings of the two contexts, six primes (3 Black; 3 White), and six words (3 pleasant; 3 unpleasant) used for the test trials were completely counterbalanced, and the order of trials was randomized.

Following the completion of this phase, participants were told that they would complete another face recognition task. This was to bolster the cover story that we were interested in their ability to remember the faces presented during the priming task. Participants were presented with 12 faces (6 targets; 6 foils) and were asked to indicate whether the face was old (by pressing the "I" key) or new (by pressing the "E" key).

Following the face recognition task, participants were told that the experiment was finished, but that we would like them to answer some questions that we were pretesting for use in future experiments. These instructions were provided to disguise the fact that the questions and the priming task were

related. Participants then responded to the 17 statements that make up the Dunton and Fazio (1997) Motivation to Control Prejudiced Responses scale. The statements were presented on a bipolar scale ranging from 1 (strongly disagree) to 7 (strongly agree). After completing this measure, participants were debriefed and thanked for their time.

Results

Behavioral effects

Trials in which errors were made (4.4%) were excluded from analyses of the priming effects. Following Maddux et al. (2005), outliers (11.2%) were defined as responses faster than 300 ms or slower than 1000 ms, and also were excluded. Responses faster than 300 ms indicate that participants responded more quickly than is considered physically possible (indicating a blind guess), whereas responses slower than 1000 ms suggest temporary attention lapses.

Facilitation scores were computed for each participant by taking the average of the two baseline response times for each word and then subtracting the corresponding response latencies of the critical fourth phase of the experiment from these baseline averages. Thus, higher scores represent greater facilitation relative to baseline. A total of eight facilitation scores were calculated for each participant, one for each of the eight cells of the 2 × 2 × 2 (Context × Race × Word Valence) within-participants factors of the design. Split half reliabilities were calculated for each of the eight facilitation scores; all *r*s > .67.

To investigate the effects of context on facilitation scores and the potential role of motivation as a moderator of these effects, we conducted an analysis using a general linear model, with Context, Race, and Word Valence entered as within-subjects dichotomous variables and MCPR as a between-subjects continuous variable (Judd, McClelland, & Smith, 1996). This analysis revealed a significant four-way interaction involving MCPR, Context, Race, and Word Valence, $F(1, 47) = 6.72$, $p = .01$. Further examination showed no significant interaction in the prison context, $p < .23$, but did reveal a significant three-way interaction in the church context, $F(1, 47) = 8.66$, $p < .01$.

To further investigate the relationship between MCPR and implicit bias in each context, an index of implicit bias was generated for each participant in each context. First, negative word trials were subtracted from positive word trials for each race of prime condition (White/Black). Then, the Black difference score was subtracted from the White difference score. Thus, each index was equal to (White prime, positive word score—White prime, negative word score)—(Black prime, positive word score—Black prime, negative word score). In this way, a single score was generated to represent the race of prime × target valence effect, with negative scores indicating a pro-Black bias, zero indicating no bias in either direction, and positive scores indicating a pro-White bias (Fazio et al., 1995; Maddux et al., 2005).

An analysis using a general linear model, with Context entered as a within-subjects dichotomous variable and MCPR as a between-subjects continuous variable, revealed a main effect for context, $F(1, 47) = 5.07$, $p = .03$. Participants displayed less pro-White bias in the church ($M = –7.21$, $SD = 135.77$) than the prison context ($M = 44.55$, $SD = 108.54$). This effect was qualified by a MCPR × Context interaction, $F(1, 47) = 4.62$, $p = .04$ (see Figure 1). Further analysis showed that there was no significant relationship between MCPR and implicit bias in the prison context, $r = .07$, $p = .64$. Implicit bias in the prison context was pro-White, regardless of motivation. In contrast, in the church context, there was a significant negative relationship between MCPR and implicit bias, $r = –.28$, $p = .05$, indicating that, as motivation increased, implicit pro-White bias diminished.

Modeling effects

We used the Quad model to examine the processes responsible for these effects. The structure of the model is depicted as a processing tree in Figure 2. 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

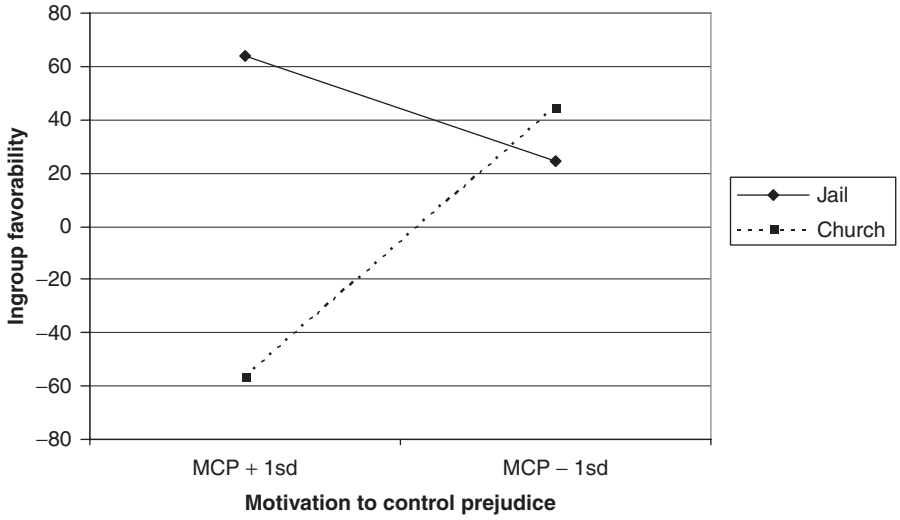

Figure 1. Implicit bias as a function of context and motivation. Negative numbers indicate a pro-Black bias. Positive numbers indicate a pro-White bias.

upon both Activation of Associations (AC) and Detection (D). Similarly, Guessing (G) is conditional upon the lack of Activation of Associations (1–AC) and the lack of Detection (1–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 positive target word following a Black face prime will be responded to correctly with the probability: AC × D × OB + (1–AC) × D + (1–AC) × (1–D) × (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 × D × OB, is the likelihood that the association between Black and negative 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) × D, is the likelihood that the association is not activated *and* that the correct response can be detected. Finally, (1–AC) × (1–D) × (G) is the likelihood that the association is not activated *and* the correct answer cannot be detected *and* that the participant guesses correctly. Furthermore, a positive word following a Black prime will be responded to incorrectly with the probability: AC × D × (1–OB) + AC × (1–D) + (1–AC) × (1–D) × (1–G. The first part of the equation, AC × D × (1–OB), is the likelihood that the association between Black and negative is activated *and* that the correct response can be detected, *but*, the association is not overcome in favor of the detected response. The second part of the equation, AC × (1–D), is the likelihood that the association is activated *and* the correct response is not detected. Finally, (1–AC) × (1–D) × (1–G) is the likelihood that the association is not activated *and* the correct response cannot be detected *and* the incorrect guess is made.

The Quad model typically estimates two AC, one OB, one D, and one G parameter for each comparison condition of interest (e.g., compatible vs. incompatible trials). However, for this study, it was necessary to alter the model because there were insufficient degrees of freedom to estimate all of these parameters. There were eight categories of responses per subject: positive and negative words following Black males in the church context, Black males in the prison context, White males in the church context, and White males in the prison context. As a result, we could estimate only eight parameters per subject. Our primary

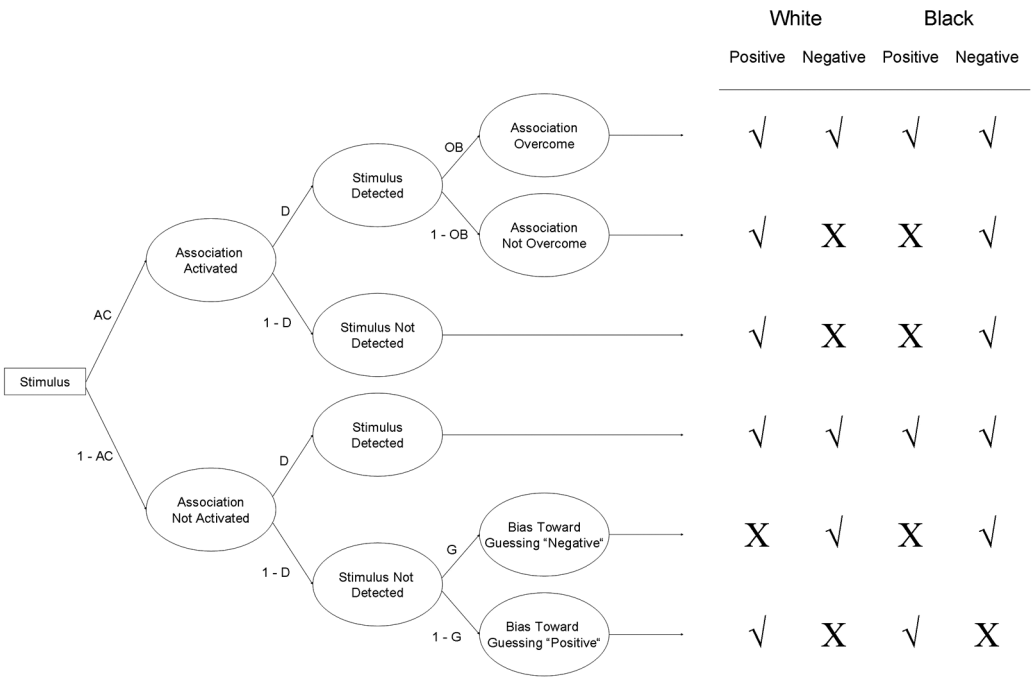

Figure 2. The Quadruple Process model of implicit task performance.

interest was to compare the AC and OB parameters for Black targets in the prison and church contexts to examine whether the reduction of bias in the church context was due to diminished activation of biased associations or enhanced inhibition of those associations. Thus, we estimated four separate AC parameters (church context Black AC, prison context Black AC, church context White AC, prison context White AC) and two separate OB parameters (a prison context OB for Black primes and a church context OB for Black primes). Finally, one D and one G parameter were estimated for each subject. These last two parameters were collapsed across all target race and context types. This procedure resulted in a saturated model for each participant, with eight degrees of freedom and eight parameters.

The respective equations for each of these eight categories were then used to predict the observed proportions of errors for each of the categories. The model's predictions were then compared to the actual data to determine the model's ability to account for the data. A χ^2-estimate was computed for the difference between the predicted and observed errors. In order to best approximate the model to the data, the parameter values were changed through maximum likelihood estimation until they produce a minimum possible value of the χ^2. The final parameter values that resulted from this process were interpreted as relative levels of the processes.

Means and standard deviations for each parameter are listed in Table 1. Individual participants' parameter estimates of Black AC and OB were submitted to separate analyses using a general linear model containing both the dichotomous within-subjects context condition (prison vs. church) and the between subjects MCPR factor (continuous variable). The Black AC parameter did not differ by context, $F(1, 47) = 0$, $p = .99$, and did not interact with MCPR, $F(1, 47) = 1.06$, $p = .31$. Analysis of the OB parameter revealed a main effect for Context, $F(1, 47) = 7.88$, $p < .01$, showing that negative associations with Black targets were overcome more effectively in the church ($M = .87$) than prison ($M = .64$)

Table 1. Mean Quad model parameter estimates and standard deviations

Parameter	Estimate	
	Prison	Church
AC		
Black–Bad	0.35 (.25)	0.34 (.37)
White–Good	0.03 (.05)	0.03 (.05)
OB		
Black–Bad	0.64 (.45)	0.87 (.33)
D	0.93 (.08)	
G	0.58 (.37)	

context. Thus, participants showed less implicit bias in the church context and were better able to overcome negative associations with Black targets in that context. Though the interaction between Context and MCPR was not statistically significant for the OB parameter, $F(1, 47) = 2.05$, $p = .16$, we tested the relationship between MCPR and OB for each of the context conditions separately in order to further examine the relationship between OB and MCPR in reducing implicit bias (see Figure 3). In the prison context, MCPR and OB were not related, $r = .07$, $p = .65$. However, in the church context, there was a marginally significant positive relationship between MCPR and OB, $r = .28$, $p = .07$. Thus, strong motivation to control prejudice was associated with reduced implicit bias in the church context and was marginally associated with successfully overcoming negative associations with Black targets in that context. Altogether, the modeling results clearly indicate that the reduction of implicit bias in the church context is associated with successful inhibition of automatic associations rather than reduced activation of those associations.[2]

Discussion

Conceptually replicating previous findings (Barden et al., 2004; Maddux et al., 2005; Wittenbrink et al., 2001), our results showed less implicit anti-Black bias when Black and White

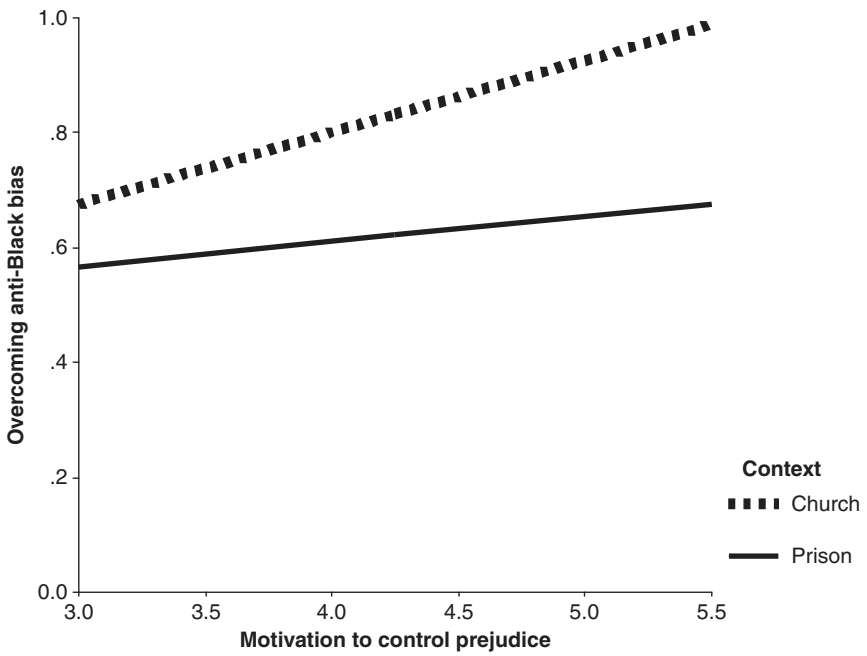

Figure 3. OB parameter estimates as a function of context and motivation.

targets were presented in positive than negative contexts. The main purpose of this experiment was to use the Quad model to shed light on the bases of this effect. Modeling analyses showed that the effect was associated with an increased ability to overcome anti-Black associations in the positive context (OB), but was not associated with changes in the nature of the anti-Black associations activated in that context (AC). This suggests that the positive context acts as a cue that facilitates the inhibition of unwanted or inappropriate associations, but does not lead to the activation of different associations, per se.

Consistent with Maddux et al. (2005), our results also showed that the context effect was moderated by participants' motivation to respond in a non-prejudiced fashion. Specifically, as motivation to respond without prejudice increased, so, too, did the tendency to show less implicit bias in the positive than negative context. Thus, highly motivated participants were more sensitive to utilizing the available cues in the environment to facilitate reduced biased responding in positive contexts. Participants with low motivation were not as sensitive to these cues and, consequently, they displayed bias regardless of the background context surrounding target primes. Modeling results provided evidence that the influence of motivation on implicit bias was associated with increased ability to overcome negative associations with Black targets. In particular, the more motivated the participants were to control prejudiced responses, the higher their OB estimates were in the church context. This suggests that highly motivated participants were especially sensitive to positive contextual cues, and that these cues triggered the overcoming bias process. The model produced no evidence that motivation was associated with the activation of different associations in the different contexts.

Though the effects of motivation on implicit bias demonstrated in this study are conceptually consistent with those of Maddux et al. (2005), they are also different in an important way. Specifically, whereas Maddux et al. (2005) found that motivation leads to the use of negative contexts as cues that reduce implicit bias, our results showed that motivation was associated with the use of positive contexts as cues that reduced bias. Other previous research has also demonstrated lower levels of implicit bias in positive than in negative contexts (e.g., Barden et al., 2004; Wittenbrink et al., 2001). Our results showed that effect, and that the effect was enhanced by a strong motivation to respond in a non-prejudiced manner. The primary difference between Maddux et al. (2005) and this study is that we presented the contexts along with the primes for brief intervals (315–545 ms), so there was very little time to process the information thoroughly. In contrast, Maddux et al. introduced contexts for longer durations prior to the presentation of primes and target words. The opportunity for more thorough processing of the contextual information could lead to greater planning for future presentations of the outgroup primes, which may be more likely with negative contexts. However, without prior context induction, those high in the motivation to control prejudice may be more sensitive to the typical valence-induced attenuation effect, and unable to formulate more deliberate strategies. Further research is called for to identify the conditions under which positive versus negative contexts are more effective in reducing bias.

Although our data suggest that inhibition is responsible for participants' ability to control implicit bias when encountering outgroup targets in positive contexts, the precise manner in which this inhibitory ability operates is an open question. We suggest three possibilities. First, the contrast in the valences of the target (e.g., Black face) and context (positive) may direct participants' awareness toward potentially undesirable negative evaluations of the target. This may enhance efforts to overcome such biases. Second, participants may be particularly likely to practice controlling unwanted biases when the targets of those biases appear in positive contexts. This may occur because unwanted biases are more likely to be detected in such contexts or because people are more likely to avoid biases in contexts that would appear to directly invalidate their implicit biases. As suggested by Monteith (e.g., Monteith et al., 2002; see also Gollwitzer & Schaal, 1998),

over time, people may begin to rely on these contexts as cues to anticipate controlling undesirable associations. Finally, a third possibility is that positive contexts facilitate inhibition by providing participants with effective distracter thoughts. According to Wegner's (1994) ironic process theory, unwanted thoughts are more easily suppressed when distracter thoughts are available that may be used to replace the unwanted thoughts. One important avenue for future research will be to examine the contributions of each of these mechanisms to individuals' effective inhibition of automatic associations.

Finally, our results have important implications for prejudice reduction strategies. We found that individuals who are highly motivated to control their bias were able to do so when contextual cues were available. Previous research has found that being motivated to control biased associations is often not enough to prevent bias from influencing responses (Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002). Devine et al. (2002) have suggested that the source of the motivation (internal vs. external) is the crucial factor in determining how efficiently people will be able to regulate bias. However, we found that highly motivated people were able to control biased responding, regardless of the source of their motivation.[3] This indicates that motivation may be effective, regardless of its source, when there are environmental cues that signal the need to regulate bias. Future research and interventions may focus on training people motivated to avoid bias to look for and rely on these cues as ways to reduce biased responses.

Notes

1. We manipulated presentation time for exploratory purposes. This manipulation did not produce any main effects or significant interactions, and will not be discussed in subsequent analyses.
2. For saturated models like the one we used (i.e., models in which the number of parameters equals the number of response categories), the value of χ^2 should equal 0. Aggregated across participants, $\chi^2 = 1.12$ in this study, suggesting that the model is capable of accounting for the behavioral data. However, statistical tests of model fit are not straightforward with saturated models. To examine the possible influence of individual variation in model fit on our findings, we calculated χ^2 values for each individual subject and entered these values as a continuous variable in our analyses. Inclusion of this variable had no influence on either the priming effects or the modeling analyses, and did not interact with any other variables.
3. The MCPR scale consists of two subscales: concern for acting prejudiced and restraint to avoid dispute. These subscales roughly correspond to internal versus external motivations for avoiding bias, respectively. We conducted analyses substituting scores for each of the subscales for the full MCPR score and found the same effects across both subscales.

Acknowledgments

This research was supported by a grant from the National Science Foundation (BCS 0820855) to the second author. We would like to thank Abdullah Dumenci, who assisted with data collection, and Paula M. Do, who assisted in preparing the stimulus materials.

References

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles of controlled and automatically activated attitudes. *Journal of Personality and Social Psychology*, *87*, 5–22.

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

Baumeister, R. F., & Vohs, K. D. (2004). *Handbook of self-regulation: Research, theory, and applications*. New York, NY: Guilford Press.

Beer, J. S., Stallen, M., Lombardo, M. V., Gonsalkorale, K., Cunningham, W. A., & Sherman, J. W. (2008). The Quadruple Process 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.

Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control-theory approach to human behavior*. New York, NY: Springer.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating

multiple processes in implicit social cognition: The quad-model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469–487.

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

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*, 316–326.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist*, *49*, 709–724.

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.

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision-making. *Journal of Personality and Social Psychology*, *91*, 797–813.

Gollwitzer, P. M., & Schaal, B. (1998). Metacognition in action: The importance of implementation intentions. *Personality and Social Psychology Review*, *2*, 124–136.

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.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.

Judd, C. M., McClelland, G. H., & Smith, E. R. (1996). Testing treatment by covariate interactions when treatment varies within subjects. *Psychological Methods*, *1*, 366–378.

Kahnenman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*, 697–720.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163–203.

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.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, *83*, 1029–1050.

Pager, D., & Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology*, *34*, 181–209.

Rokeach, M., & Ball-Rokeach, S. J. (1989). Stability and change in American value priorities, 1968–1981. *American Psychologist*, *44*, 775–784.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.

Schuman, H., Steeh, C., & Bobo, L. (1985). *Racial attitudes in America: Trends and interpretations*. Cambridge, MA: Harvard University Press.

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., 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.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Smedley, B. D., Stith, A. Y., & Nelson, A. R. (2003) *Unequal treatment: Confronting racial and ethnic disparities in health care*. Washington, DC: National Academies Press.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, *101*, 34–52.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81*, 815–827.

Biographical notes

THOMAS J. ALLEN is a Ph.D. student in the department of Psychology at the University of California, Davis. His research interests fall into three primary areas: the self-regulation of automatic responses, the effects of stereotypes on perception and memory, and how self-evaluation motives influence the expression of prejudice.

JEFFREY W. SHERMAN is a Professor in the department of Psychology at the University of California, Davis. His research investigates the cognitive processes underlying social psychology and behavior. In particular, he is interested in how stereotypes and prejudice affect how people perceive themselves, other people, and groups of people. He is currently Editor of the journal *Social Cognition*.

KARL CHRISTOPH KLAUER is a Professor in the department of Psychology at the University of Freiburg, Germany. His research investigates the cognitive processes underlying social psychology and behavior. In particular, he is interested in implicit measures of attitudes and stereotypes. He is currently Associate Editor of the journal *Social Psychological and Personality Science*.