CHAPTER 13

Self-Control Over Automatic Associations

Karen Gonsalkorale, Jeffrey W. Sherman, and Thomas J. Allen

ABSTRACT

Processes that permit control over automatic impulses are critical to a range of goal-directed behaviors. This chapter examines the role of self-control in implicit attitudes. It is widely assumed that implicit attitude measures reflect the automatic activation of stored associations, whose expression cannot be altered by controlled processes. We review research from the Quad model (Sherman et al., 2008) to highlight the importance of two self-control processes in determining the influence of automatically activated associations. The findings of this research indicate that processes relating to detecting appropriate responses and overcoming associations contribute to performance on implicit attitude measures. These two processes work together to enable self-control of automatic associations; one process detects that control is needed, and the other process overcomes the associations to permit correct behavior. Implications for understanding self-control dilemmas are discussed.

Keywords: Implicit attitudes, automatic associations, self-regulation, detection, overcoming bias

It is Friday morning and John is excited because he is finally going on the road trip that he has been planning for months. He loads his bags into the car, completes a safety check, reviews the road atlas for the umpteenth time, and pulls out of the driveway. With his favorite music blaring and the traffic flowing smoothly, John is feeling happy and relaxed. A few minutes later, he makes a wrong turn and finds himself heading towards the office instead of his intended destination. Realizing his mistake, John curses himself for being on "auto-pilot."

In this example, John's automatic habit has prevented him from successfully executing his trip (at least temporarily). What did John need to do to achieve his goal? First, he had to be able to identify the correct route to his vacation. Having virtually committed the route to memory, John more than satisfied this first condition. Given that he knew the required route, John then needed to overcome the automatic response that compelled him to follow his commute to work. On this second count, John failed.

We argue that these processes—detecting appropriate responses and regulating automatic habits—are critically important across a wide range of goal-directed behaviors. Our approach is not limited to instances in which self-control resolves conflicts between lower-level and higher-order goals (e.g., satisfying a sugar craving versus staying on a diet). Rather, we propose that self-control processes also will be crucial whenever a situation is characterized by competition between automatic impulses

and responses that promote goal attainment. In the example above, John has no goal to go to work; he simply has an automatic habit[1] that temporarily disrupts his vacation plans. Many psychological phenomena have the same basic structure. For example, in the Stroop task (Stroop, 1935), the automatic habit to read the word must be overcome to report the color of the word accurately. In implicit measures of attitudes, automatic associations with targets (e.g., associations between Blacks and guns) must be overridden to perform the task accurately when the task requires an association-incompatible response. Though participants generally seek to perform these tasks correctly, rarely do they have a goal to implement habitual responses or activate automatic associations in the course of performing them. Thus, although the automatic and controlled processes produce competing responses, this conflict typically does not arise from competing goals.

Contribution to Understanding Self-Control Dilemmas

We explore self-control issues by investigating the processes that contribute to performance on tasks that place automatic and controlled processes in opposition to one another. Specifically, in our research, we have applied the Quadruple Process Model (Quad model; Conrey et al., 2005; Sherman et al., 2008) to dissociate the processes that influence responses on such tasks. As described below, this model assesses the extent to which individuals detect correct responses and overcome automatic associations. The model also estimates the degree to which automatic associations are activated while performing the task and the influence of response biases. On the Stroop task, for example, the Quad model is able to assess the relative influence of processes relating to the automatic habit to read, accuracy in identifying the color of the words, regulation of the automatic habit, and response biases, as contributors to task performance.

Our primary level of analysis is the mind. We are interested in understanding the cognitive, affective, and motivational processes that enable individuals to control their attitudes and behavior in social life. A primary goal of our research is to separate the multiple automatic and controlled processes that underlie people's attitudes and behaviors. This approach has important linkages to levels of analysis at the brain and at society. Given that automaticity and control are associated with distinct regions in the brain, a complete account of the processes involved in self-control requires consideration of the neural systems that underlie them. Conversely, neuroscientific approaches may benefit from mapping a range of automatic and controlled processes (observed behaviorally) onto specific brain systems. Our level of analysis also has implications for societal-level analyses. The automatic and controlled processes occurring within individuals may influence the effectiveness of society's efforts at controlling its citizens. For instance, anti-discrimination laws may only be effective if individuals are willing and able to control their automatic stereotypes and prejudices. Reverse effects may occur also; society may facilitate or constrain the extent to which automatic and controlled processes affect thought and behavior. If stereotypes are salient within a society, for example, individuals may require stronger self-control to combat the effects of the stereotypic associations that they hold. The relationships between the three levels of analysis highlight the importance of a multi-disciplinary approach to the issue of self-control.

Although our approach is applicable to a wide range of situations, in this chapter we will review research findings in the domain of implicit attitudes,[2] which has been the primary focus of our work thus far. It is widely assumed that implicit attitude measures reflect the unintended, automatic activation of stored associations, whose expression cannot be altered or inhibited by controlled processes (e.g., Bargh, 1999; Devine, 1989; Fazio et al., 1995; Greenwald, McGhee, & Schwartz, 1998). Thus, self-control issues have been seen as largely irrelevant to understanding responses on implicit measures. In contrast, we propose that both automatic and controlled processes underlie implicit task performance, and that these processes can be independently measured using the Quad model.

Consider the Stroop task again. 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, to perform the task accurately, the automatic habit to read the word must be overcome 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, many of which have the same compatibility structure as the Stroop task. The identical responses of two individuals on an Implicit Association Test (IAT; Greenwald et al., 1998), for example, may reflect moderately biased associations (e.g., between Blacks and negativity) in one case, but strong associations that are successfully overcome in the other.

Thus, investigating self-control within implicit attitude measures is important for gaining a more complete understanding of what these measures assess and how they should be conceptualized. For example, common interpretations of implicit measures may underestimate not only the extent of controlled processing, but also the extent of automatic processing because a strong ability to overcome automatic bias may mask the true extent of that bias. Another implication is that self-control in implicit task performance may be partly responsible for a host of effects that are typically attributed to the operation of automatic processes. Findings that scores on implicit attitude measures vary across individuals and are responsive to experimental interventions (*see* Blair, 2002) have often been interpreted as evidence that automatic associations differ among individuals and can be readily changed. However, if controlled processes also are responsible for variability and malleability in implicit task performance, then the implications of such results would be very different. For example, the results may not indicate the ease with which implicit associations may be changed but rather may reflect a greater role for intentions and motivations than has been previously assumed. As a final example, better understanding the role of self-control in implicit attitudes may help to better identify means for changing those attitudes. If an implicit bias stems from biased automatic associations, then a strategy that directly influences those associations may be most effective. In contrast, if the bias stems from deficits of self-control, then interventions that improve self-control may be most effective.

Importantly, the implications of our research extend well beyond the exertion of self-control during implicit task performance. Because implicit measures and tasks like the Stroop create self-control needs that mirror those encountered in everyday life, exploring the processes required to successfully perform these tasks can enhance understanding of many real-world self-control dilemmas. In particular, this approach can shed light on any situation in which a goal may be thwarted in favor of an automatic response. Although we do not view conflict between lower-level and higher-order goals as a necessary feature of self-control dilemmas, our approach may nevertheless yield insight into such situations. The ability to recruit self-control processes to perform a task effectively will likely predict success at mediating between immediate and longer-term goals. For example, how well a person who wants to quit smoking is able to overcome positive associations with cigarettes on an implicit attitude measure might predict whether she later smokes a cigarette to satisfy her nicotine craving (lower goal), or behaves in line with her desire to stop smoking (higher goal). Thus, our work contributes to understanding how individuals behave in situations that require control over impulses, both at the task level and at the broader goal level.

Control and Automaticity in Social Behavior

The Quad model was developed, in part, by considering the processes that appear across a wide spectrum of dual-process models of social and cognitive psychology (e.g., Chaiken & Trope, 1999; Sherman, 2006). By definition, all four processes of the Quad model are

never found within any particular dual-process model. Rather, the goal of dual-process models is to assess the extent to which a judgment or behavior reflects one type of automatic processing and one type of controlled processing. The Quad model incorporates the processes that are most commonly identified across the various dual-process models. These processes have been shown to be fundamental and ubiquitous components of judgment and behavior.

Although they are not always explicitly presented as such, dual-process models almost always are relevant to questions of self-control. They are concerned with delineating the circumstances under which judgment and behavior are driven by controlled, intended processes versus automatic, unintended processes. In examining these questions, dual-process models have generally been concerned with one of two different types of control. In some models, control is characterized by stimulus detection processes that attempt to provide an accurate depiction of the environment. For example, in dual-process models of persuasion, the controlled process is involved in discrimination between strong and weak arguments (e.g., Chaiken, 1980; Petty & Cacioppo, 1981; Fazio, 1990). In models of impression formation, the controlled process entails attention to and integration of target behaviors, providing an individuated (and presumably accurate) impression of the person (e.g., Brewer, 1988; Fiske & Neuberg, 1990). In Jacoby's Process Dissociation models (Jacoby, 1991; Lindsay & Jacoby, 1994; Payne, 2001), control represents an ability to determine and provide a correct response.

However, in other dual-process models, control is characterized by self-regulatory processes that attempt to inhibit unwanted or inappropriate information. For example, in Devine's (1989) model of stereotyping, control must be exerted to overcome the automatic influence of stereotypes. In Wegner's (1994) model of thought suppression, control must be exerted to inhibit unwanted thoughts. In many models of social judgment, self-regulatory control is exerted when people try to correct their judgments for subjectively expected biases (e.g., Martin, 1986; Wegener & Petty, 1997). These types of dual-process models have been more explicitly recognized as pertaining to self-control.

Both detection and regulation processes are controlled processes in that they require intention and cognitive resources, and can be terminated at will (e.g., Bargh, 1994). However, they have different functions and have very different influences on attitudes and behavior. It is clear that, on many occasions, they operate simultaneously. For example, a police officer's decision whether to shoot a Black man who may or may not have a gun may depend both on his ability to discriminate whether the man has a gun and, when there is no gun, his ability to overcome an automatic bias to associate Blacks with guns and to shoot. Thus, we believe that there is much to be gained by distinguishing between these types of control and measuring their contributions to behavior independently.

Dual-process models also have generally been concerned with one of two different types of automaticity. Most commonly, automaticity is represented as simple associations that are triggered by the environment without the perceiver's awareness or intent. Stereotypes play this role in dual-process models of impression formation (e.g., Brewer, 1988; Fiske & Neuberg, 1990). In models of persuasion (e.g., Chaiken, 1980; Petty & Cacioppo, 1981) and judgment (e.g., Epstein, 1991; Sloman, 1996), heuristics function in much the same way. This is the kind of automaticity that implicit attitude measures are intended to assess (e.g., Devine, 1989; Fazio et al., 1995; Greenwald et al., 1998).

In other dual-process models, however, automatic processes influence behavior only when control fails. For example, Jacoby's Process Dissociation model of memory (Jacoby, 1991) proposes that, when controlled attempts at recollection fail, people may instead rely on automatically generated feelings of familiarity to identify the stimulus as old. Others have portrayed the influence of implicit stereotypes in (mis)identifying weapons as operating in this manner (e.g., Payne, 2001). Another example is the implicit preference shown for items on the right side of a display when conscious introspection provides no rational basis for this preference (Nisbett & Wilson, 1977).

Although both types of automatic processes may operate without conscious intention, awareness, or the use of cognitive resources, clearly they are different kinds of processes. For example, a police officer's decision to shoot might be influenced by automatically activated associations between Blacks and guns. In the absence of such associations, however, the officer's decision might still be influenced by a secondary automatic bias to presume danger in the absence of clear evidence to the contrary, and guess that the person is holding a gun. We believe it is important to distinguish between these types of automatic processes and to measure their contributions to behavior separately.

The Quad Model

The Quad model is a multinomial model (*see* Batchelder & Riefer, 1999) designed to estimate the independent contributions of each of the four processes described above to a given behavior. 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 discriminates between appropriate and inappropriate responses. Sometimes, the activated associations conflict with the detected correct response. For example, on incompatible trials of the Stroop task or incompatible trials of implicit attitude measures (e.g., a Black face prime followed by a positive target word), automatic associations or habits 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. Guessing can be random, but it may also reflect a systematic tendency to prefer a particular response. For example, pressing the "unpleasant" key in response to a target face in the IAT (Greenwald et al., 1998) 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. Thus, guessing can be relatively automatic or controlled. The Quad model employs multinomial modeling to estimate the influence of each of these processes within a single task (for a review, *see* Batchelder & Riefer, 1999).

The structure of the Quad model is depicted as a processing tree in Figure 13–1. 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 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). Note that these conditional relationships do not imply a serial order in the onset and conclusion of the different processes. Rather, these relationships are mathematical descriptions of the manner in which the parameters interact to produce behavior. Thus, attempts to detect a correct response (D) and attempts to overcome automatic biases (OB) may occur simultaneously. However, in determining a response on a trial of a given task, the influence of attempts to overcome bias will be seen only in cases in which detection is successful.

The conditional relationships described by the model form a system of equations that predict the number of correct and incorrect responses in different conditions (e.g., compatible and incompatible trials). We will illustrate with reference to the Black-White IAT (Greenwald et al., 1998), one of the most frequently used implicit measures of attitudes toward Blacks and Whites. On each trial of the IAT, participants are presented in the middle of

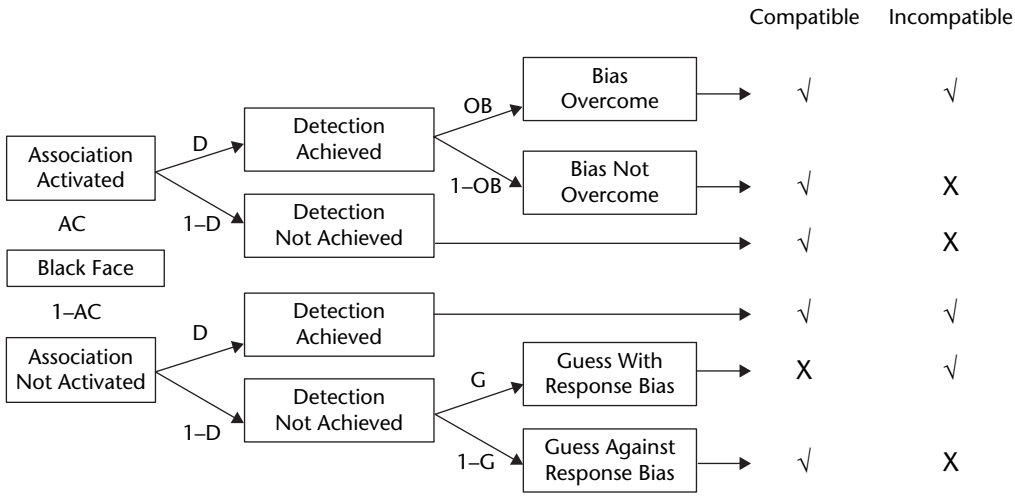

Figure 13–1. The Quadruple Process Model (Quad Model). Each path represents a likelihood. Parameters with lines leading to them are conditional upon all preceding parameters. The table on the right side of the figure depicts correct (√) and incorrect (X) responses as a function of process pattern and trial type. In this particular figure, the response bias refers to guessing with the positive key.

a computer screen with a stimulus from one of four categories: Black faces, White faces, pleasant words, and unpleasant words. Participants are asked to indicate, as quickly and accurately as possible, to which category the stimulus belongs by pressing the appropriate key, according to labels at the top of the screen. In the "compatible" block, participants are instructed to press one key in response to Black faces and unpleasant words, and the other key in response to White faces and pleasant words. The keys used to categorize Black and White faces are switched in the "incompatible" block, such that the Black and pleasant categories are assigned to one key, and the White and unpleasant categories to the other key. Participants who respond more quickly in the compatible block compared to the incompatible block are thought to have implicit negative associations toward Blacks relative to Whites.

According to the Quad model, a Black face stimulus in an incompatible block of a Black–White IAT will be assigned to the correct side of the screen 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 is activated *and* that the correct answer can be detected *and* that the association is overcome in favor of the detected response. The second part of the equation, $(1 - AC) \times D$, is the likelihood that the association is not activated *and* that the correct response can be detected. Finally, $(1 - AC) \times (1 - D) \times G$, is the likelihood that the association is not activated *and* 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 face stimulus will return the correct answer. The respective equations for each item category (e.g., Black faces, White faces, positive words, and negative words in both compatible and incompatible blocks) are then used to predict the observed proportion 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 $\chi 2$-estimate is computed for the difference between the predicted and observed errors. To best approximate the model to the data, the four 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. For a complete description of data analysis within the Quad model, *see* Conrey et al. (2005).

Behavioral and Neurological Evidence for the Validity of the Quad Model's Parameters

Numerous research findings have established the validity of the Quad model. Conrey et al. (2005) demonstrated that the Quad model accurately describes performance on two of the most widely used implicit measures of attitudes, the IAT and the sequential priming task (e.g., Payne, 2001). Because Detection and Overcoming Bias are relatively controlled processes, fewer cognitive resources should decrease the extent of D and OB. One experiment showed that limiting participants' ability to engage in controlled processing by forcing them to respond quickly reduced estimates for D and OB, but did not affect Activation of Associations or Guessing. Thus, capacity constraints influence only the relatively controlled processes. In another experiment, manipulating the base rate of correct responses requiring a right- or left-handed button press shifted response bias (G) in the direction of the base rate, but did not affect any of the other parameters. These findings indicate that the four processes of the Quad model contribute to performance on implicit attitude measures, respond to experimental manipulations in a predictable manner, and are empirically separable.

Other research findings provide further support for the construct validity of the Quad model's parameters. For example, reaction time bias on the IAT is positively correlated with AC, but negatively correlated with OB (Conrey et al., 2005). This finding suggests that greater association activation increases implicit racial bias, but the ability to inhibit automatic associations attenuates this bias. Furthermore, a neuro-imaging study of IAT performance (Beer et al., 2008) showed that AC was correlated with activity in the amygdala and insula, which are involved in emotional processing and arousal (Phan et al., 2006; Phelps et al., 2000). This finding is consistent with the depiction of AC as measuring association activation. At the same time, on trials in which automatic associations and controlled processes compete to determine performance (i.e., incompatible trials), D was associated with activation in both the dorsal anterior cingulate cortex (dACC) and the dorsolateral prefrontal cortex (DLPFC). Whereas activity in the dACC has been related to detecting conflict between competing behavioral responses (e.g., Botvinick et al., 1999), activity in the DLPFC has been linked to inhibitory control over pre-potent responses (e.g., Chee et al., 2000; Taylor et al., 1998). Thus, when automatic and controlled processes compete to direct behavior, the D parameter predicts brain activity associated with detecting appropriate behavior among competing responses and inhibiting inappropriate automatic reactions. This is consistent with the Quad model's depiction of D as a controlled process that selects appropriate behavior and feeds into efforts to overcome inappropriate automatic influences.[3]

The next section reviews research in which we applied the Quad model to examine the processes underlying implicit attitudes. This research illustrates the benefits of considering self-control issues when exploring implicit attitudes.

Applying the Quad Model: Detection and Overcoming Bias in Implicit Attitudes

We have conducted a series of studies to demonstrate that performance on implicit attitude measures can vary and be changed through a number of different mechanisms. In these studies, we used the Quad model to analyze new and published data on implicit attitude variability and malleability. These findings highlight the utility of using the Quad model to gain deeper insight into the different processes responsible for implicit task performance. As we will describe below, some of these insights were masked in previous research by data analytic approaches that could not separate the influences of Detection, Overcoming Bias, and Association Activation.

Public versus Private Contexts

In one of the earliest demonstrations of how the Quad model can shed new light on existing data, Conrey et al. (2005) re-analyzed a study by Lambert et al. (2003) on the effects of public versus private contexts on implicit attitudes. Lambert et al. (2003) found that bias on the Weapons Identification Task (Payne, 2001) was amplified when participants believed that their performance would be made public. Two competing explanations have been advanced for this effect. According to the habit-strengthening account, public contexts lead to enhanced bias by increasing the influence of dominant responses. In contrast, the impairment of control account proposes that anticipation of a public setting creates a cognitive load that weakens the ability to control responses.

Conrey et al.'s (2005) Quad model re-analysis of the Lambert et al. (2003) data indicated that the public context led to an increase in Activation of Associations and a decrease in Detection. These findings are consistent with both the habit strengthening and the impairment of control explanations. Moreover, the Quad model analysis revealed an increase in Overcoming Bias in the public context, a finding that was not predicted by either account, and could not be detected in the original analyses based on Process Dissociation (Jacoby, 1991). This finding indicates that public accountability does not impair all aspects of self-control. Although it does reduce the ability to determine correct responses, it also increases people's ability to overcome unwanted bias. More broadly, this finding is important because it shows the value of separating different types of control that may be influenced in opposite ways by the same context.

Individual Differences in Motivation to Respond Without Prejudice

We also have applied the Quad model to understand motivation-based individual differences in the expression of racial bias. Research suggests that individuals who are either internally or externally motivated to respond without prejudice show lower levels of bias on explicit measures compared to individuals who are not motivated to control prejudice (Plant & Devine, 1998). However, only those individuals who are internally but not externally motivated (high IMS/low EMS participants) are able to respond without bias on implicit prejudice measures (Amodio, Devine, & Harmon-Jones, 2008; Amodio et al., 2003; Devine et al., 2002). By contrast, individuals motivated by internal and external reasons (high IMS/high EMS participants) and those who are not internally motivated (low IMS participants) exhibit bias on implicit measures.

Although there is substantial evidence that high IMS/low EMS individuals are effective in regulating race bias on implicit measures, relatively little is known about how they achieve non-prejudiced responding. Recently, Amodio et al. (2008) found that high IMS/low EMS individuals showed heightened electrophysiological responses corresponding to conflict detection following stereotypical errors on a priming measure of implicit stereotypes. Hence, high IMS/low EMS participants showed heightened conflict monitoring when their responses were discrepant with the goal to be non-prejudiced. This suggests that these individuals are particularly adept at detecting competing appropriate and inappropriate responses. If this is the case, then these same individuals should demonstrate higher estimates of the Quad model's Detection parameter than other participants. To test this hypothesis, we (Sherman et al., 2008) re-analyzed the accuracy data from Amodio et al. (2008).

In addition to analyzing the Detection parameter, we examined the possibility that high IMS/low EMS individuals also have less biased associations automatically activated and/or greater ability to overcome those associations than other individuals. Just as enhanced detection would produce less implicit racial bias among high IMS/low EMS individuals, so, too, would reduced activation of biased associations or a greater ability to regulate those associations. Based on an application of Jacoby's (1991) PD procedure, Amodio et al. (2008) reported no differences in automatic activations among individuals with different motivations.

Our Quad model re-analysis of the data indicated that, compared to the other participants, the high IMS/low EMS participants were more able to detect appropriate and inappropriate responses on the weapons identification task. In addition, these individuals exhibited less activation of stereotypic associations compared to individuals who were not internally motivated to respond without prejudice. We replicated these findings in a follow-up study using a different measure of implicit attitudes. Specifically, we found that high IMS/low EMS participants showed less implicit racial bias on the IAT (Greenwald et al., 1998), higher estimates of Detection, and lower levels of Association Activation, compared to other individuals. Importantly, there was no evidence of differences in Overcoming Bias as a function of different motivations in either study.

These findings are consistent with Amodio et al.'s conclusion that high IMS/low EMS participants are more effective in controlling race bias because they have a more finely tuned conflict detection system. For conflict detection to occur, the correct and incorrect responses must be identified. Our results also suggest that high IMS/low EMS individuals may have less biased automatic associations to overcome compared to other individuals. The finding that the OB parameter did not differentiate among individuals with different motivations would appear to indicate that motivation-based individual differences in implicit bias may have little to do with inhibition processes. As such, it seems that high IMS/low EMS individuals may not be especially good at regulating, per se. Rather, it appears that they are particularly able to detect when regulation is required (i.e., when there are conflicting responses and a danger of responding inappropriately; Amodio et al., 2008; Monteith et al., 2002), thereby increasing the likelihood of behaving appropriately.

Training to Negate Biased Associations

The findings described in the previous section may shed light on how certain prejudice reduction strategies work. If high IMS/low EMS individuals show less implicit bias because of enhanced behavioral monitoring and lower association activation, then prejudice reduction strategies that improve awareness of appropriate and inappropriate responses should also reduce implicit bias via the same mechanisms. To test this possibility, we trained participants to negate anti-Black and pro-White associations before performing the IAT (Sherman et al., 2008). On each trial of the training task a Black or White face was presented together with a positive or negative word. Participants in the negate associations condition were instructed to press the "YES" key whenever they saw a Black face with a positive word below it or a White face with a negative word below it, and to press the "NO" key whenever a Black face appeared with a negative word or a White face appeared with a positive word. Participants in the maintain associations condition were given the opposite instructions. This type of training has been shown to reduce implicit stereotyping in previous research (Kawakami et al., 2000). We hypothesized that the attention and effort required to successfully execute the training task would enhance Detection and reduce Association Activation. Based on our finding that high IMS/low EMS individuals did not show greater levels of Overcoming Bias compared to other individuals, we did not expect the training to improve OB.

Results supported our hypotheses. Replicating Kawakami et al.'s (2000) findings, the participants who had been trained to negate negative associations with Blacks and positive associations with Whites showed significantly less IAT bias than the participants who were trained to maintain these associations. Analysis using the Quad model showed that the negation training not only weakened participants' automatically activated associations (AC) but also improved their ability to determine the correct response (D). The finding of enhanced detection is consistent with the idea that training enables individuals to develop cues for recognizing and then controlling non-prejudiced responses (Monteith et al., 2002). This suggests that people can be trained to engage self-control in a manner similar to individuals who are internally motivated to be non-prejudiced (and, presumably, train themselves).

In other studies, we have identified important roles for Overcoming Bias in accounting for variability in implicit attitudes. As a measure of the extent to which activated associations can be overcome when they conflict with appropriate behavior, we reasoned that OB should be particularly sensitive to variations in self-regulatory ability. To test this hypothesis, we investigated populations that are known to differ in self-regulatory ability in two studies. The first study examined the processes underlying the effects of alcohol intoxication on implicit racial bias.

Alcohol Intoxication

Research indicates that alcohol impairs cognitive and motor performance by reducing the ability to regulate prepotent responses. For example, intoxicated individuals are less able to inhibit distracting thoughts and restrain inappropriate responses on cognitive tasks (Easdon & Vogel-Sprott, 2000). Applying these findings to the domain of social attitudes, Bartholow, Dickter, and Sestir (2006) hypothesized that alcohol increases stereotypic responding by impairing self-regulatory ability. To explore this possibility, they modified a priming measure of implicit racial stereotyping (e.g., Dovidio, Evans, & Tyler, 1986) such that it included "go" trials and "stop" trials. The primes were Black or White faces and houses (the control primes), and the target words consisted of adjectives that can be used to describe people or houses (e.g., *carpeted, furnished*). Half of the person adjectives were stereotypic of Blacks (e.g., *athletic*, *lazy*) and half were stereotypic of Whites (e.g., *intelligent, boring*). For the "go" trials, participants were instructed to indicate whether the adjective could ever be used to describe the picture of the person or house that preceded it. Participants were instructed to withhold responses on the "stop" trials. Errors on the stop trials served as the behavioral index of regulation failure (Logan & Cowan, 1984). Results indicated that alcohol affected the pattern of errors on the stop trials, such that the high-dose group committed significantly more errors on stereotype-consistent trials than on stereotype-inconsistent trials, whereas the placebo group did not show this stereotyping effect. Moreover, electrophysiological data indicated that the alcohol-based impairment of inhibition of stereotype-based responses was mediated by a reduction in the negative slow wave (NSW) component of the event-related potential. The NSW has been associated with the engagement of cognitive control processes sub-serving inhibition (West & Alain, 1999). At the same time, alcohol had no effect on a neurological marker of stereotype activation (the P300). These behavioral and neurological data suggest that alcohol impairs the ability to regulate the expression of stereotypic associations.

Applying the Quad model to Bartholow et al.'s priming data, we found that Overcoming Bias was the only parameter that differed across alcohol consumption conditions (Sherman et al., 2008). This finding suggests that alcohol intoxication interferes with people's ability to regulate automatically activated associations. There were no significant differences in Association Activation or Detection as a function of alcohol consumption. Thus, the Quad model findings provide converging evidence that the effects of alcohol on race-based responding are specific to inhibition failure. The Quad model offers a non-invasive means of detecting such effects with behavioral data alone.

Aging

In our second study that focused on the Overcoming Bias parameter, we explored the effects of aging on implicit racial bias (Gonsalkorale, Sherman, & Klauer, 2009). Large national surveys have consistently shown that older White Americans tend to express more racial prejudice than their younger counterparts (e.g., Wilson, 1996). Recent research suggests that age-related differences in racial bias extend to the implicit level, with one large study (Nosek, Banaji, & Greenwald, 2002) reporting a positive correlation between age and implicit racial bias. These findings often are interpreted as evidence that older people's racial associations are more biased than those of younger adults, reflecting generational changes in societal attitudes.

An alternative explanation for age differences in prejudice is that deficits in self-regulatory ability alter the attitudinal expression of older adults. Given that the ability to inhibit automatically activated stereotypes enables people to behave non-prejudicially (Bartholow et al., 2006; Devine, 1989; Moskowitz, Salomon, & Taylor, 1999), and that inhibitory functioning declines with age (Connelly, Hasher, & Zacks, 1991; Hasher & Zacks, 1988), losses in inhibitory ability may increase stereotyping and prejudice during old age, even if the underlying associations are of equivalent (or even declining) strength across the life span. Consistent with this possibility, von Hippel, Silver, and Lynch (2000) found that losses in inhibitory functioning mediated increases in explicit racial stereotyping among the elderly. This research also indicated that, contrary to popular wisdom, older adults reported strong desires to control their prejudices, suggesting that they were willing but not able to control their biases.

We conducted a study to examine whether inhibitory processes can account for age differences in racial bias on an implicit measure. Race IAT data were collected from White participants who visited the IAT demonstration Web site (http://implicit.harvard.edu/; Nosek et al., 2002). We modeled the data as a function of participant age, which ranged from 11 to 94. The results suggested that age-related differences in IAT bias arose from differences in the ability of older and younger adults to regulate automatically activated associations. Despite showing stronger IAT effects, the older adults demonstrated less activation of biased associations and a greater likelihood of detecting the correct response than the younger adults. However, as predicted, Overcoming Bias decreased with age. It appears that, despite weaker activation of associations and greater detection of correct responses, the older adults exhibited stronger implicit bias behaviorally because they were less able to inhibit their activated associations. These findings suggest that age differences in implicit racial bias may be caused by age-related losses in regulatory functions.

Predicting the Quality of Intergroup Interactions

The findings reviewed above indicate that Detection and Overcoming Bias are important underlying processing components of performance on implicit measures of racial bias. We believe that the processes that direct performance on these immediate response tasks are likely to predict success at resolving impulse regulation conflicts in broader domain-relevant contexts. To illustrate this relationship, we will now describe a study in which we examined the ability of the Quad model's parameters to predict behavior in an intergroup interaction (Gonsalkorale, von Hippel, Sherman, & Klauer, 2009).

The goal of this study was to test hypotheses regarding the processes underlying the relationship between implicit attitudes and behavior in intergroup interactions. According to one account, implicit race bias predicts unfriendly behavior in cross-race interactions because biased automatic associations drive prejudice-consistent behavior in attitude-relevant situations (e.g., Dovidio, Kawakami, & Gaertner, 2002). In previous research, correlations between scores on implicit attitude measures and interaction behavior (e.g., Dovidio et al., 2002; Dovidio et al., 1997; Fazio et al., 2005; McConnell & Leibold, 2001) have been taken as evidence that automatic associations direct behavior. However, implicit attitude measures are not pure reflections of the automatic associations that are hypothesized to drive behavior in intergroup settings (Amodio et al., 2004; Bartholow et al., 2006; Conrey et al., 2005; Payne, 2001; Sherman, 2009; Sherman et al., 2008). Thus, correlations between scores on these measures and behavior in the presence of outgroup members do not necessarily indicate the influence of those associations. In contrast, we tested this idea directly by examining whether Activation of Associations predicts poorer interaction quality.

An alternative account proposes that people may be able to prevent their automatic biases from influencing behavior by regulating their behavior when interacting with outgroup members (e.g., Richeson & Shelton, 2007). Consistent

with this possibility, one study (Shelton et al., 2005) found that 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. In our study, we wanted to examine whether the immediate regulation of automatic associations, as reflected in responses on implicit measures, is sufficient to influence interaction behavior. 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. If overcoming associations contributes to smooth intergroup interactions, Overcoming Bias should predict better interaction quality.

To examine these issues, we asked non-Muslim Caucasian participants to interact with an experimental confederate who appeared to be and was described as Muslim. Following the interaction, the confederate rated how much he liked the participants, whereas the participants completed a Go/No-Go Task (GNAT; Nosek & Banaji, 2001) measuring implicit bias toward Muslims. The GNAT is a variant of the IAT that assesses attitudes toward a single target group (e.g., Muslims) rather than relative evaluations of two groups (e.g., Muslims versus non-Muslims). Participants who showed more negative attitudes toward Muslims on the GNAT were evaluated less positively by the Muslim confederate. We applied the Quad model to the GNAT data to examine the extent to which different processes may contribute to the quality of the interaction. The confederate's ratings of how much he liked the participants were predicted by an interaction between automatic negative associations and ability to overcome bias. 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. Thus, among participants with strong anti-Muslim associations, the ability to recruit self-control to perform the GNAT effectively predicted success at regulating behavior during the intergroup interaction. These findings are the first to show that process estimates derived from the Quad model are related to self-control in a broader behavioral context that plays out over extended time.

This study also illustrates how the Quad model may enhance interpretability of data from implicit attitude measures. We found that participants who exhibited greater bias against 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 modelling 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 strengths of the Quad model.

Summary and Implications

Our findings highlight the importance of self-control processes in determining the influence of automatically activated associations. Across multiple studies, we have found that processes relating to both detecting appropriate responses and overcoming associations contribute to performance on measures of implicit knowledge. These two processes work together to permit

self-control of automatic associations; one process detects that control is needed, and the other process overcomes the associations to permit correct behavior.

The idea that different types of controlled processes play a role in implicit task performance has important implications for understanding the nature of implicit attitudes. If implicit measures are presumed to reflect only the automatic activation of associations, then malleable performance on such measures must, by definition, be taken as evidence that the associations activated in performing a given task have been altered. That is, implicit attitude malleability must reflect either changes in the nature of the underlying associations or changes in the particular associations that are temporarily accessible. However, our research suggests that such conclusions are likely to significantly overestimate the extent to which activated associations can be altered. In some cases, malleability effects will be due, at least in part, and maybe entirely, to response processes that have nothing to do with the underlying associations, per se. As such, though implicit attitude malleability is certainly cause for optimism about people's ability to avoid automatic stereotyping and prejudice effects (e.g., Blair, 2002), caution is warranted in concluding that associative knowledge is easily altered.

Implications for Treating Implicit Measures as "Process Pure"

There is now considerable evidence that implicit measures engage multiple processes, both automatic and controlled, as underlying associations are translated into behavioral responses on the tasks. Our research shows the consequences of this task complexity for interpreting implicit attitude effects. Specifically, our findings highlight the danger of assuming a one-to-one correspondence between performance on implicit measures and the extent of automatic association activation. It follows that implicit measures should not be assumed to provide estimates of processes (e.g., Fazio et al., 1995), representations (Greenwald et al., 1998; Wilson et al., 2000), or systems (Rydell & McConnell, 2006; Strack & Deutch, 2004) that are independent from intention and control. In the studies described earlier, we demonstrated these conclusions in the domains of implicit attitude variability and malleability. However, our findings have implications for other implicit attitude effects, as well. For example, dissociations between implicit and explicit measures cannot be assumed to reflect the separate and independent contributions of automatic and controlled processes, representations, or systems to performance on the two tasks. One difficulty for dual representation and dual system models is the frequent lack of correlations among different implicit measures of the same attitude (e.g., Sherman, 2006). If all implicit measures are tapping the same automatic process, representation, or system, then different implicit measures should correlate more highly than they often do. However, from the current perspective, performance on the different measures may reflect a variety of differences in the processes recruited in performing the tasks. The key differences between any two measures (implicit or explicit) may have to do with the nature of the associations activated by the different tasks, the nature of response biases in performing the tasks, or the nature of more controlled detection or self-regulation processes engaged while performing the tasks. Finally, the same considerations surround interpretations of the relationships between implicit measures and behavior. It is not necessarily the case that correlations between an implicit measure and a behavior reflect the operation of automatic processes, representations, or systems; other components of task performance may also (or instead) be responsible for the relationship.

Implications for Attitude Change Strategies

An important question in the minds of people who are interested in promoting egalitarianism is how to design successful prejudice-reduction strategies. If interventions designed to change implicit attitudes do not always lead to changes in the activation of associations, then how useful are they? In our view, changing underlying

associations is but one method of changing people's implicit attitudinal and behavioral responses. Our research suggests that different attitude-change strategies may be best suited to changing different kinds of implicit attitudes. For example, if the attitudinal bias stems not from biased associations but from an inability to monitor ongoing behavior for appropriateness (as among high IMS/high EMS participants), then interventions designed to enhance the detection of conflicting responses might be advised. However, if the attitudinal bias stems from a deficit in self-regulation (as among older adults), then the best intervention might be one that serves to strengthen the ability to overcome unwanted associations. The current research demonstrates how such an attitude-intervention matching process may be achieved by identifying the bases for individual differences in implicit attitudes and the bases of the effects of interventions on implicit attitude change.

Future Directions: Predicting Success in Achieving Broader Goals

At the beginning of this chapter, we argued that application of the Quad model to immediate response tasks (e.g., implicit measures) holds promise for increasing understanding of broader goal-directed behavior. The ability to recruit and apply self-control processes on immediate response tasks is likely to predict success at resolving conflicts within the same domain between low-level, narrow goals and high-level, global goals. The findings from the intergroup interaction study provide the first indication of such a relationship between task-level control and the regulation of broader domain-relevant behaviors and goals.

Our ongoing research aims to provide further evidence for this application of the model by examining other relationships between process estimates derived from task performance and real-world behavior. For example, research has shown that smokers have less negative implicit attitudes about cigarettes than do ex-smokers and non-smokers (Sherman et al., 2003). We are applying the Quad model to try to understand the reasons for this effect, and to help generate effective interventions to help people quit smoking. For example, it may be that smokers have less negative automatic associations with cigarettes than do non-smokers. Alternatively, it may be that smokers are less able to determine appropriate smoking-related behaviors than non-smokers, or are less able to regulate the expression of their more favorable associations. By understanding how these groups differ on these processes, we can better understand why some people start smoking and others do not, why some people are able to quit smoking and others cannot, and what specific processes might need to be addressed in interventions aimed at reducing smoking.

As a general model of impulse control, the Quad model is relevant to a range of self-control dilemmas that are characterized by competing goals. Thus, the model may be able to predict whether dieters will choose healthy foods in the wake of tempting alternatives, whether recovering gambling addicts will be enticed by the lure of a casino, when people will be able to control affective reactions such as anger or happiness that may interfere with important decisions, and so on. In these scenarios and many others, automatic response tendencies that satisfy lower goals also have the potential to thwart higher order goals in a manner described by the Quad model. It is our hope that the model's broad applicability will lead to enhanced understanding of self-control issues in many different domains of judgment and behavior.

Notes

1 Note that this habit could reflect automatic goal activation occurring at an unconscious level. Our main point here is that the approach described in this chapter need not involve the activation or application of goals (either consciously or unconsciously).

2 In this chapter we focus on applying the Quad model to implicit measures, which we argue reflect the joint operation of automatic associations and controlled processes. However, there are many instances in which the association, impulse, or habit may not be automatic, per se.

The Quad model is relevant to these situations, as long as the association or impulse conflicts with a desired or intended response (thus producing incompatible responses).

3 For methodological reasons having to do with the different trials used to derive estimates of overcoming bias (OB) and brain activity, we were unable to associate that parameter with specific brain activity.

References

Amodio, D. M., Devine, P. D., & Harmon-Jones, E. Individual differences in the regulation of intergroup bias: The role of conflict monitoring and neural signals for control. J Pers Soc Psychol 2008; 94: 60–74.

Amodio, D. M., Harmon-Jones, E., & Devine, P. G. Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. J Pers Soc Psychol 2003; 84: 738–753.

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. Neural signals for the detection of unintentional race bias. Psychol Sci 2004; 15: 88–93.

Bargh, J. A. The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In: Wyer, R. S., & Srull, T. K. (Eds.), *Handbook of social cognition, Vol. 1: Basic processes*, 2nd ed. Hillsdale, NJ: Erlbaum, 1994: pp. 1–40.

Bartholow, B. D., Dickter, C. L., & Sestir, M. A. Stereotype activation and control of race bias: Cognitive control of inhibition and its impairment by alcohol. J Pers Soc Psychol 2006; 90: 272–287.

Batchelder, W. H., & Riefer, D. M. Theoretical and empirical review of multinomial process tree modeling. Psychon Bull Rev 1999; 6: 57–86.

Beer, J. S., Stallen, M., Lombardo, M. V., Gonsalkorale, K., Cunningham, W. A., & Sherman, J. W. The Quadruple Process Model approach to examining the neural underpinnings of prejudice.NeuroImage 2008; 42: 775–783.

Beck, A. T. *Cognitive therapy and the emotional disorders*. New York, NY: International Universities Press, 1976.

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature 1999; 402: 179–181.

Brewer, M. B. A dual process model of impression formation. In Srull, T. K., & Wyer, R. S. (Eds.), *Advances in social cognition*, Vol. 1. Hillsdale, NJ: Erlbaum, 1988: pp. 1–36.

Chaiken, S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. J Pers Soc Psychol 1980; 39: 752–766.

Chaiken, S., & Trope, Y. (Eds.). *Dual-process theories in social psychology*. New York, NY: Guilford Press, 1999.

Chee, M. W. L., Sriram, N., Soon, C. S., & Lee, K. M. Dorsolateral prefrontal cortex and the implicit association of concepts and attributes. *Neuroreport* 2000; 11; 135–140.

Connelly, S. L., Hasher, L., & Zacks, R. T. Age and reading: The impact of distraction. Psychol Aging 1991; 6: 533–541.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. Separating multiple processes in implicit social cognition: The Quad Model of implicit task performance. J Pers Soc Psychol 2005; 89: 469–487.

Devine, P. G. Stereotypes and prejudice: Their automatic and controlled components. J Pers Soc Psychol 1989; 56: 5–18.

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. J Pers Soc Psychol 2002; 82: 835–848.

Dovidio, J. F., Evans, N., & Tyler, R. B. Racial stereotypes: The contents of their cognitive representations. J Exp Soc Psychol 1986; 22: 22–37.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. Implicit and explicit prejudice and interracial interaction. J Pers Soc Psychol 2002; 82: 62–68.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. On the nature of prejudice: Automatic and controlled processes. J Exp Soc Psychol 1997; 33: 510–540.

Easdon, C. M., & Vogel-Sprott, M. Alcohol and behavioral control: Impaired response inhibition and flexibility in social drinkers. Exp ClinPsychopharm 2000; 8: 387–394.

Epstein, S. Cognitive-experiential self theory: An integrative theory of personality. In Curtis, R. (Ed.), *The self with others: Convergences in psychoanalytical, social, and personality psychology*. New York, NY: Guilford, 1991: pp. 111–137.

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. Aging and prejudice: Diminished regulation of

automatic race bias among older adults. J Exp Soc Psychol 2009; 45: 410–414.

Gonsalkorale, K., von Hippel, W., Sherman, J. W., & Klauer, K. C. Bias and regulation of bias in intergroup interactions: Implicit attitudes toward Muslims and interaction quality. J Exp Soc Psychol 2009; 45: 161–166.

Fazio, R. H. Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. Adv Exp Soc Psychol 1990; 23: 75–109.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. Variability in automatic activation as an unobstrusive measure of racial attitudes: A bona fide pipeline? J Pers Soc Psychol 1995; 69: 1013–1027.

Fiske, S. T., & Neuberg, S. L. A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. Adv Exp Soc Psychol 1990; 23: 1–74.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. Measuring individual differences in implicit cognition: The Implicit Association Test. J Pers Soc Psychol 1998; 74: 1464–1480.

Hasher, L., & Zacks, R. T. Working memory, comprehension, and aging: A review and a new view. In Bower, G. H. (Ed.), *The psychology of learning and motivation: Advances in research and theory*, Vol. 22. San Diego, CA: Academic Press, 1998: pp. 193–225.

Jacoby, L. L. A process dissociation framework: Separating automatic from intentional uses of memory. J Mem Lang 1991; 30: 513–541.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. J Pers Soc Psychol 2000; 78: 871–888.

Lambert, A. J., Payne, B., Jacoby, L. L., Shaffer, L. M., Chasteen, A. L., & Khan, S. R. Stereotypes as dominant responses: On the "social facilitation" of prejudice in anticipated public contexts. J Pers Soc Psychol 2003; 84: 277–295.

Lindsay, D. S., & Jacoby, L. L. Stroop process-dissociations: The relationship between facilitation and interference. J Exp Psychol Hum Percept Perform 1994; 20: 219–234.

Logan, G. D., & Cowan, W. B. On the ability to inhibit thought and action: A theory of an act of control. Psychol Rev 1984; 91: 295–327.

Martin, L. L. Set/reset: Use and disuse of concepts in impression formation. J Pers Soc Psychol 1986; 51: 493–504.

Miller E. K., & Cohen J. D. An integrative theory of prefrontal cortex function. Ann Rev Neurosci 2001; 24: 167–202.

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. Putting the brakes on prejudice: On the development and operation of cues for control. J Pers Soc Psychol 2002; 83: 1029–1050.

Moskowitz, G. B., Salomon, A. R., & Taylor, C. M. Preconsciously controlling stereotyping: Implicitly activated egalitarian goals prevent the activation of stereotypes. Soc Cogn 2000; 18: 151–177.

Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. Psychol Rev 1977; 84: 231–259.

Nosek, B. A., Banaji, M., & Greenwald, A. G. Harvesting implicit group attitudes and beliefs from a demonstration web site. Group Dynam Theory Res Pract 2002; 6: 101–115.

Payne, B. K. Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. J Pers Soc Psychol 2001; 81: 181–192.

Payne, B. K., Lambert, A. J., & Jacoby, L. L. Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. J Exp Soc Psychol 2002; 38: 384–396.

Petty, R. E., & Cacioppo, J. T. *Attitudes and persuasion: Classic and contemporary approaches*. Dubuque, IA: Brown, 1981.

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. Neuroimage 2002; 16: 331–348.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., et al. Performance on indirect measures of race evaluation predicts amygdala activation. J Cogn Neurosci 2000; 12: 729–738.

Plant, E. A., & Devine, P. G. Internal and external motivation to respond without prejudice. J Pers Soc Psychol 1998; 75: 811–832.

Richeson, J. A., & Shelton, J. N. Negotiating interracial interactions: Costs, consequences, and possibilities. Curr Direct Psychol Sci 2007; 16: 316–320.

Rydell, R. J., & McConnell, A. R. Understanding implicit and explicit attitude change: A systems

of reasoning analysis. J Pers Soc Psychol 2006; 91: 995–1008.

Shelton, J. N., Richeson, J. A., Salvatore, J., & Trawalter, S. Ironic effects of racial bias during interracial interactions. Psychol Sci 2005; 16: 397–402.

Sherman, J. W. On building a better process model: It's not only how many, but which ones and by which means? Psychol Inq 2006; 17: 173–184.

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

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. The self-regulation of automatic associations and behavioral impulses. Psychol Rev 2008; 115: 314–335.

Sherman, S. J., Rose, J. S., Koch, K., Presson, C. C., & Chassin, L. Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. J Soc Clin Psychol 2003; 22: 13–39.

Sloman, S. A. The empirical case for two systems of reasoning. Psychol Bull 1996; 119: 3–22.

Strack, F., & Deutsch, R. Reflective and impulsive determinants of human behavior. Pers Soc Psychol Rev 2004; 8: 220–247.

Stroop, J. R. Studies of interference in serial verbal reactions. J Exp Psychol 1935; 18: 643–662.

Taylor, S. F., Kornblum, S., Lauber, E. J., Minoshima, S., & Koeppe, R. A. Isolation of specific interference processing in the Stroop Task: PET Activation studies. J Neuroimaging 1998; 6: 81–92.

von Hippel, W., Silver, L. A., & Lynch, M. E. Stereotyping against your will: The role of inhibitory ability in stereotyping and prejudice among the elderly. Pers Soc Psychol Bull 2000; 26: 523–532.

Wegener, D. T., & Petty, R. E. The flexible correction model: The role of naïve theories of bias in bias correction. Adv Exp Soc Psychol 1997; 29: 141–208.

Wegner, D. M. Ironic processes in mental control. Psychol Rev 1994; 101: 34–52.

West, R., & Alain, C. Event-related neural activity associated with the Stroop task. Cogn Brain Res 1999; 8: 157–164.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. A model of dual attitudes. Psychol Rev 2000; 107: 101–126.