

A University of California author or department has made this article openly available. Thanks to the Academic Senate's Open Access Policy, a great many UC-authored scholarly publications will now be freely available on this site.

Let us know how this access is important for you. We want to hear your story!

[http://escholarship.org/reader\\_feedback.html](http://escholarship.org/reader_feedback.html)



## Peer Reviewed

### Title:

The Self-Regulation of Automatic Associations and Behavioral Impulses

### Journal Issue:

Psychological Review, 115(2)

### Author:

[Sherman, JW](#)  
[Gawronski, B](#)  
[Gonsalkorale, K](#)  
[Hugenberg, K](#)  
[Allen, TJ](#)  
[Groom, CJ](#)

### Publication Date:

04-01-2008

### Series:

[UC Davis Previously Published Works](#)

### Permalink:

<http://escholarship.org/uc/item/4x59z6nr>

### DOI:

<https://doi.org/10.1037/0033-295X.115.2.314>

### Local Identifier(s):

UCPMS ID: 1510210

### Abstract:

The distinction between automatic processes and controlled processes is a central organizational theme across areas of psychology. However, this dichotomy conceals important differences among qualitatively different processes that independently contribute to ongoing behavior. The Quadruple process model is a multinomial model that provides quantitative estimates of 4 distinct processes in a single task: the likelihood that an automatic response tendency is activated; the likelihood that a contextually appropriate response can be determined; the likelihood that



eScholarship  
University of California

eScholarship provides open access, scholarly publishing services to the University of California and delivers a dynamic research platform to scholars worldwide.

automatic response tendencies are overcome when necessary; and the likelihood that in the absence of other information, behavior is driven by a general response bias. The model integrates dual-process models from many domains of inquiry and offers a generalized, more nuanced framework of impulse regulation across these domains. The model offers insights into many central questions surrounding the operation and the interaction of automatic and controlled processes. Applications of the model to empirical and theoretical concerns in a variety of areas of psychology are discussed. © 2008 American Psychological Association.

**Copyright Information:**

All rights reserved unless otherwise indicated. Contact the author or original publisher for any necessary permissions. eScholarship is not the copyright owner for deposited works. Learn more at [http://www.escholarship.org/help\\_copyright.html#reuse](http://www.escholarship.org/help_copyright.html#reuse)



**eScholarship**  
University of California

eScholarship provides open access, scholarly publishing services to the University of California and delivers a dynamic research platform to scholars worldwide.

# The Self-Regulation of Automatic Associations and Behavioral Impulses

Jeffrey W. Sherman  
University of California, Davis

Bertram Gawronski  
University of Western Ontario

Karen Gonsalkorale  
University of California, Davis

Kurt Hugenberg  
Miami University

Thomas J. Allen  
University of California, Davis

Carla J. Groom  
KRC Research

The distinction between automatic processes and controlled processes is a central organizational theme across areas of psychology. However, this dichotomy conceals important differences among qualitatively different processes that independently contribute to ongoing behavior. The Quadruple process model is a multinomial model that provides quantitative estimates of 4 distinct processes in a single task: the likelihood that an automatic response tendency is activated; the likelihood that a contextually appropriate response can be determined; the likelihood that automatic response tendencies are overcome when necessary; and the likelihood that in the absence of other information, behavior is driven by a general response bias. The model integrates dual-process models from many domains of inquiry and offers a generalized, more nuanced framework of impulse regulation across these domains. The model offers insights into many central questions surrounding the operation and the interaction of automatic and controlled processes. Applications of the model to empirical and theoretical concerns in a variety of areas of psychology are discussed.

*Keywords:* automaticity, cognitive control, dual-process models, impulsive behavior, self-regulation

Behavior is often influenced by mental associations, habits, feelings, and impulses that are activated automatically, without intention or awareness, and that may be difficult to control. Sometimes, these automatic processes interfere with our ability to behave in a desired or appropriate fashion. For example, automatically activated associations between Black Americans and aggression may interfere with a police officer's ability to refrain from shooting an unarmed Black man (Correll, Park, Judd, & Wittenbrink, 2002). Likewise, the automatized habit of driving on the right side of the road may hamper one's attempts to navigate traffic in Great Britain. As a more mundane

example, in the Stroop task (Stroop, 1935), the automatic habit to read the word *blue* must be overcome to report the color of the word accurately (if, say, the color of the ink is red).

The purpose of this article is to describe a model that accounts for how automatically activated mental constructs and competing self-regulatory processes interact to direct behavior. The model specifies the likelihood of different outcomes and, more important, provides a mathematical means to estimate the independent contributions of different processes in producing those outcomes. Thus, the model estimates the influences both of the automatic associations (or feelings, impulses, or habits) that affect behavior and of the controlled processes that may work in opposition to the associations to achieve effective self-control. At the same time, the model also details the manner in which these different components constrain one another in producing behavior.

A central feature of the model is that it goes beyond the basic division between automatic processes and controlled processes. We argue that there are important distinctions among the types of processes that have typically been characterized as automatic and controlled and that these distinctions can provide a more detailed and nuanced description of human behavior than can standard dual-process depictions of automaticity versus control. In so doing, the model integrates a great variety of dual-process models from many domains of inquiry and offers a generalized framework for understanding impulse and impulse control across these domains.

For this purpose, we first review common conceptualizations of automaticity and control and argue that the common distinction between automatic processes and controlled processes conceals

---

Jeffrey W. Sherman, Karen Gonsalkorale, and Thomas J. Allen, Department of Psychology, University of California, Davis; Bertram Gawronski, Department of Psychology, University of Western Ontario, London, Ontario, Canada; Kurt Hugenberg, Department of Psychology, Miami University; Carla J. Groom, KRC Research, London, England.

This research was supported by NIMH Grant MH59774 to Jeffrey W. Sherman and by grants from the Canada Research Chairs Program (CRC) and the Academic Development Fund of the University of Western Ontario (ADF) to Bertram Gawronski.

We are grateful to Rainer Banse, Wilhelm Hofmann, Jacob Matthews, Thorsten Meiser, Cindy Pickett, and Jim Sherman for helpful comments on an earlier version of this article. We extend special thanks to Riki Conrey, who played a pivotal role in the development of this research program and contributed to the content of this article.

Correspondence concerning this article should be addressed to Jeffrey W. Sherman, Department of Psychology, University of California, Davis, 1 Shields Avenue, Davis, CA 95616. E-mail: jsherman@ucdavis.edu

important differences between qualitatively distinct processes. We then outline the basic assumptions of our Quadruple process model (Quad model; Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005) and describe the application of the model to a number of important theoretical issues surrounding automaticity and control. Finally, we describe how this model relates to current models of self-regulation and to different dual-process theories of automatic and controlled processing. We argue that the Quad model can elucidate research on automaticity and on its regulation across many areas of psychology. An initial description and test of the Quad model was presented by Conrey et al. (2005). The current article significantly expands the theoretical elaboration of the model, provides comparisons with related models, and demonstrates the application of the model to a wide variety of research questions and domains of interest.

### MULTIPLE AUTOMATIC AND CONTROLLED PROCESSES

Automatic processes often are defined as those that occur without intention or awareness, that require few cognitive resources to enact, and that are difficult to terminate once they have been initiated. In contrast, controlled processes are those that typically require intention and cognitive resources. People can initiate and halt these processes at will, and people are usually aware of their operation (Bargh, 1994; Moors & De Houwer, 2006; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). This distinction between automatic processes and controlled processes now occupies a central role in many areas of psychology, including research on visual attention (e.g., Egeth & Yantis, 1997), learning (e.g., Reber, 1993), memory (e.g., Jacoby, 1991), reasoning and problem solving (e.g., Epstein, 1994; Sloman, 1996), judgment and decision making (e.g., Ferreira, Garcia-Marques, Sherman, & Sherman, 2006; Kahneman, 2003), affect and emotion (e.g., Lazarus, 1982; Zajonc, 1980), anxiety and depression (e.g., Mathews & MacLeod, 1994), text processing and comprehension (e.g., McKoon & Ratcliff, 1992), attributional processes (e.g., Gilbert, 1989; Trope, 1986), impression formation (e.g., Brewer, 1988; Fiske & Neuberg, 1990), stereotyping and prejudice (e.g., Banaji & Greenwald, 1995; Devine, 1989), and persuasion (e.g., Chen & Chaiken, 1999; Petty & Wegener, 1999). Across these areas, there has been a proliferation of so-called dual-process models that are aimed at describing the contributions of automatic and controlled processes to behavior (Chaiken & Trope, 1999). However, although the distinction between automatic processing modes and controlled processing modes is ubiquitous in psychology, different dual-process models emphasize different types of automatic and controlled processes. Thus, the distinction takes on different meanings in different contexts.<sup>1</sup>

#### Conceptualizations of Control

In many dual-process theories, the controlled process is one that seeks out, attends to, or extracts information from the environment to provide accurate and appropriate judgments and behaviors. In the distinction between explicit memory and implicit memory, for instance, control is exerted in explicit memory search to retrieve previously stored information (Roediger, 1990). In dual-process models of judgment and decision making, control is exerted to reason the correct answer from available information (e.g., Epstein, 1994; Ferreira et al., 2006; Kahneman, 2003; Sloman, 1996).

In dual-process models of persuasion (Chen & Chaiken, 1999; Petty & Wegener, 1999), control is exerted to weigh the strengths and weaknesses of a persuasive message.

However, in other dual-process theories, the controlled process is self-regulatory in nature. Here, control is exerted to inhibit or suppress inappropriate, unwanted, or incorrect responses (e.g., Baumeister & Vohs, 2004; Carver & Scheier, 1981). For example, in Wegner's (1994) model of thought suppression, control must be exerted to inhibit unwanted thoughts that spring to mind automatically. Similarly, in Devine's (1989) dual-process model of stereotyping, it is proposed that controlled, self-regulatory processing is necessary to overcome automatically activated stereotypes.

Contemporary dual-process models tend to focus on either one or the other of these two processes. However, though the two processes may be similar in that both require intention and cognitive resources and both can be initiated and terminated at will, it is clear that control exerted toward accuracy and control exerted toward suppression are very different types of processes. Consistent with this assumption, several studies have demonstrated behavioral and neurological dissociations between these processes (e.g., Beer, Shimamura, & Knight, 2004; Jonides et al., 2003; Miller & Cohen, 2001; Miyake et al., 2000). It also is 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 the color of the ink (accuracy), and they must suppress the habit of reading the word (suppression). Similarly, a police officer's decision as to whether to fire his weapon at a Black man holding an ambiguous object (see Correll et al., 2002; Greenwald, Oakes, & Hoffman, 2003; Payne, 2001) depends on both the ability to identify whether the man has a gun (accuracy) and, if he has no gun, the ability to overcome a stereotypic bias to associate Black people with guns and to shoot (suppression). It is important to note that even though both processes have been described as controlled, they may be affected in opposite ways by the same context or manipulation, as we demonstrate below. Thus, a fuller appreciation of the complexity of these behaviors would consider and measure both processes simultaneously.

#### Conceptualizations of Automaticity

Automatic processes have also been conceptualized in two different ways in dual-process models. Most commonly, automatic processes are depicted as simple associations or habitual responses that are triggered by environmental stimuli without the perceiver's awareness or intent, capturing attention and drawing it away from more deliberate processes (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. This kind of automaticity also describes response interference mechanisms in sequential priming phenomena. Zajonc (1980) showed that objects are processed affectively before any controlled processing is engaged. This automatic activation of affect can interfere with controlled processes, such that

<sup>1</sup> The distinction between automatic processes and controlled processes is not always the defining feature of the alternate processes in various dual-process models. Nevertheless, these alternate processes do possess features of automatic versus controlled processes.

the affective reactions elicited by a prime stimulus can either facilitate or inhibit evaluative responses to a target stimulus (Fazio, Sanbonmatsu, Powell, & Kardes, 1986; for a review, see Klauer & Musch, 2003). Such interference effects form the basis for many implicit measures of attitudes, such as affective priming (Fazio, Jackson, Dunton, & Williams, 1995), the Implicit Association Test (IAT) (Greenwald, McGhee, & Schwartz, 1998), and the Extrinsic Affective Simon Task (De Houwer, 2003).

In other dual-process models, however, automatic processes play a different role. Rather than interfering with controlled processing and capturing behavior at the outset of a response, automatic processes may instead act as a secondary source of responding that direct behavior only when control fails. In this case, intentional, controlled processes first constrain the response options that come to mind, and automatic processes operate only when the early selection fails to provide an adequate response. Jacoby's model (e.g., Jacoby, 1991; Jacoby, Toth, & Yonelinas, 1993) of recognition memory is perhaps the most prominent early selection model. Correct recognition of old items on a memory test can be achieved through controlled memory search or through the use of feelings of familiarity automatically generated by old items at test (Jacoby, 1991; Mandler, 1980). According to Jacoby (1991; Jacoby et al., 2005, 1993), people attempt first to consciously recollect the details of the test items. Only if this controlled process fails to produce recollection do feelings of familiarity guide responses (see also Bowles et al., 2007; Jacoby, Shimizu, Daniels, & Rhodes, 2005). Another example of such a secondary bias is the widespread preference for items placed on the right side of a display when conscious introspection provides no rational basis for a particular preference (Nisbett & Wilson, 1977).

Extant dual-process models typically focus on one or the other process, but not both. However, though they may be similar in that they both tend to operate without intention, awareness, or the use of cognitive resources, it is clear that automatic processes that are engaged only in the wake of failed control and automatic processes that capture attention at the outset of a response are quite different. For example, Wagner and his colleagues (Wagner, Stebbins, Masciari, Fleischman, & Gabrieli, 1998) demonstrated a dissociation between the brain regions associated with repetition priming effects and effects of familiarity in recognition memory. It also is clear that many behaviors may be simultaneously influenced by both processes. For instance, a police officer's split-second decision to fire his weapon at a Black suspect holding an ambiguous object might be influenced by automatically activated associations between Black men and guns (see Correll et al., 2002; Greenwald et al., 2003). In the absence of such associations, however, the officer's decision might still be influenced by a general tendency to presume danger in the absence of clear evidence to the contrary (e.g., Conrey et al., 2005; Jacoby, Bishara, Hessels, & Toth, 2005). Thus, a fuller appreciation of the complexity of such behaviors would consider and measure both processes simultaneously.

## THE QUAD MODEL

The four processes described above show up repeatedly in dual-process models across many areas of psychology. 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 of behavioral response

(Conrey et al., 2005). In the model, it is proposed that many behaviors depend jointly on the activation of an impulsive response tendency (activation [AC]), the ability to determine a contextually appropriate or a correct response (detection [D]), the success at overcoming impulsive response tendencies, when necessary (overcoming bias [OB]), and the influence of general guessing or response biases that may influence behavior in the absence of other available guides to response (guessing [G]).<sup>2</sup> Thus, the Quad model synthesizes all of the four processes that have previously been discussed for dual-process models. One of the most important features of the Quad model, however, is that it goes beyond a mere narrative description of the proposed processes. Rather, the basic assumptions of the Quad model have been implemented as a multinomial model (Batchelder & Riefer, 1999; Riefer & Batchelder, 1988) that can provide independent quantitative estimates of each of the four processes proposed in the model.<sup>3</sup>

## Components of the Model

### Activation

AC represents the likelihood that an association, evaluation, or behavioral impulse is activated upon encountering a stimulus object. The stronger the link to the stimulus object, the more likely associations, evaluations, or behavioral impulses are to be activated by the stimulus. AC resembles the initial, stimulus-driven, attention capture type of process implied in many dual-process theories. Given the automatic nature of AC, estimates of AC should be relatively independent of awareness, intention, and availability of cognitive resources.

### Detection

D represents the likelihood of determining a contextually appropriate or a correct response in a given task or behavior. D represents the type of accuracy-oriented process found in dual-process models of recollection memory (e.g., Jacoby, 1991; Roediger, 1990), judgment and decision making (e.g., Epstein, 1994; Sloman, 1996), or persuasion (e.g., Chen & Chaiken, 1999; Petty & Wegener, 1999). In the Quad model, D represents the likelihood that the correct response can be determined rather than the likelihood that the correct response is provided. In a Stroop task, for example, detection (i.e., identification of color) should be fairly high for all (color-seeing) individuals regardless of their scores on the actual task. Occasional interference from the word-reading habit does not negate the fact that given the opportunity, most adults can determine the color of the ink. Given that accurate detection reflects a largely controlled process,

<sup>2</sup> Note that the current designations of the AC and the D parameters as *activation* and *detection* deviates from our earlier designations as *association activation* and *discriminability* (see Conrey et al., 2005). The use of the current terminology is based on the broader meaning of the two parameters in the context of self-regulation, which goes beyond earlier applications of the Quad model.

<sup>3</sup> In no way do we intend to imply that the four processes represented in the Quad model are exhaustive. To be sure, many other processes may be important in different contexts (see Sherman, 2006b). However, the four parameters of the Quad model are the processes that show up most frequently in dual-process models, and the four parameters have been shown to be relevant across many domains of inquiry.



D should be influenced by intentions and cognitive capacity. Greater motivation to perform well should increase D, whereas fewer cognitive resources should reduce D. Theoretically, however, through extensive practice in a given context, D may be routinized, and take on features of more automatic processes.

### *Overcoming Bias*

OB represents the likelihood that an activated association or behavioral impulse was overcome and was replaced by a contextually appropriate or accurate response. Thus, OB is relevant only when an activated bias (AC) is incompatible with an appropriate response determined via detection (D). For example, on a congruent trial in the Stroop task (e.g., the word *red* written in red ink), both word reading (AC) and identification of the color (D) lead to the same (correct) response (i.e., red). However, on an incongruent trial (e.g., the word *red* written in blue ink), only detection of the color (D) leads to the correct response (i.e., blue), whereas word reading (AC) leads to an incorrect response (i.e., red). In terms of the Quad model, OB determines whether AC or D directs behavior. If the bias is overcome, then D will drive the response. However, if OB fails, then AC drives the response. As such, OB resembles the type of self-regulatory process found in Wegner's (1994) model of thought suppression or Devine's (1989) model of stereotype control. Given that suppression represents a controlled process, OB, like D, should be influenced by both cognitive capacity and motivation. Fewer available resources should make overcoming bias more difficult, but greater motivation to succeed at self-regulation should increase OB. As with D, theoretically, OB could be automatized through extensive practice.

### *Guessing*

When no association or impulse is activated and a correct response cannot be determined, then a guess (G) must be made. Thus, G is a bias that operates only in the absence of all other bases of responding. G encompasses secondary influences of responding, similar to the familiarity component of Jacoby's (1991) model of process dissociation (PD). Guessing need not be random but may instead reflect a systematic bias to prefer a particular response. It is important to note that guessing also need not be an automatic process. Indeed, it is easy to imagine cases in which people adopt conscious guessing strategies to guide their responses on a given task. For example, if the incorrect categorization of an object as negative could be considered a socially undesirable response (e.g., in a measure designed to assess implicit prejudice), participants may adopt a conscious guessing strategy to respond "positive" rather than "negative" (e.g., Conrey et al., 2005). In other cases, however, G may reflect a largely unconscious process, such as the aforementioned tendency to prefer objects on the right side of a display (Nisbett & Wilson, 1977).

### *The Structure of the Model*

A useful example to illustrate the general structure of the Quad model is the so-called shooter bias (e.g., Correll et al., 2002; Greenwald et al., 2003; Payne, 2001). The term *shooter bias* is used to describe the enhanced likelihood of pulling the trigger of a gun in response to Black suspects, even when the suspect holds a harmless object rather than a dangerous weapon. To investigate the shooter bias in the laboratory,

Payne (2001) developed a paradigm that resembles a standard sequential priming task (for alternative tasks, see Correll et al., 2002; Greenwald et al., 2003). In this so-called weapon identification task, participants are presented with face primes of either Black individuals or White individuals for 200 ms. The face prime is then replaced by a target picture of either a gun or a tool, which remains on the screen for another 200 ms. Immediately afterward, a black-and-white pattern mask appears on the screen. Participants' task is to indicate as quickly as possible whether the presented target object was a gun or a tool.

From the perspective of the Quad model, performance in the weapon identification task (and in real-life situations resembling the task) depends on the following: (a) the activation of stereotypical associations between Black people and guns or White people and tools, which may elicit a race-related tendency to respond "gun" for Black individuals or "tool" for White individuals (AC); (b) the general ability to identify the presented object as a gun or tool (D); (c) the success at replacing race-related response tendencies with the appropriate response implied by the accurate detection of the object when the response tendency is incompatible with the identified correct response (OB); and (d) a general response bias to guess either "gun" or "tool" when there is no other source of information (G).

The interplay of these processes and their respective outcomes are depicted as a processing tree in Figure 1. In the tree, each path represents the likelihood of a correct response. Processing parameters with lines leading to them are conditional on all preceding parameters. For instance, overcoming bias (OB) is conditional on both association activation (AC) and detection (D). Similarly, guessing (G) is conditional on the lack of association activation ( $1 - AC$ ) and lack of detection ( $1 - D$ ). Note that these conditional relationships do not imply a temporal sequence in the onset and conclusion of the different processes. Rather, these relationships are strictly mathematical descriptions of the manner in which the parameters interact to produce behavior. Thus, determining the correct response (D) and overcoming activated associations (OB) may occur simultaneously. However, in determining a response on a trial of a given task, the influence of overcoming bias (OB vs.  $1 - OB$ ) will be seen only in cases in which associations are activated (AC) and detection is successful (D).<sup>4</sup>

<sup>4</sup> Theoretically, the model assumes that all four processes may be initiated simultaneously and interact in an ongoing fashion. However, the mathematical implementation of the model can only reveal the actions of the different processes in conditional circumstances. Both theoretically and mathematically, the activation (AC) and the detection (D) parameters are interchangeable. There is no conditionality between these two parameters. In contrast, the mathematical implementation dictates that efforts at overcoming bias (OB) will be seen only on incompatible trials and only in cases in which AC and D both have occurred. Similarly, the model reveals influences of guessing biases (G) only when AC and D have failed. Thus, although, theoretically, OB and G may occur from the moment of stimulus onset, the model is only equipped to detect the influence of these processes in particular conditional cases. If one wished to examine, for instance, self-regulatory processes that occur from the onset of stimulus presentation, regardless of the occurrence of AC or D, then one would need to construct a different tree model with different conditional relationships among the parameters. Note that although the ability to measure OB is dependent on successful AC and D, the specific level of OB reported by the model is independent of the specific likelihoods of AC and D, assuming each is greater than zero. That is, given that the parameter estimates are constrained to be greater than zero and less than one, the probability estimates of each of the parameters are independent of one another.

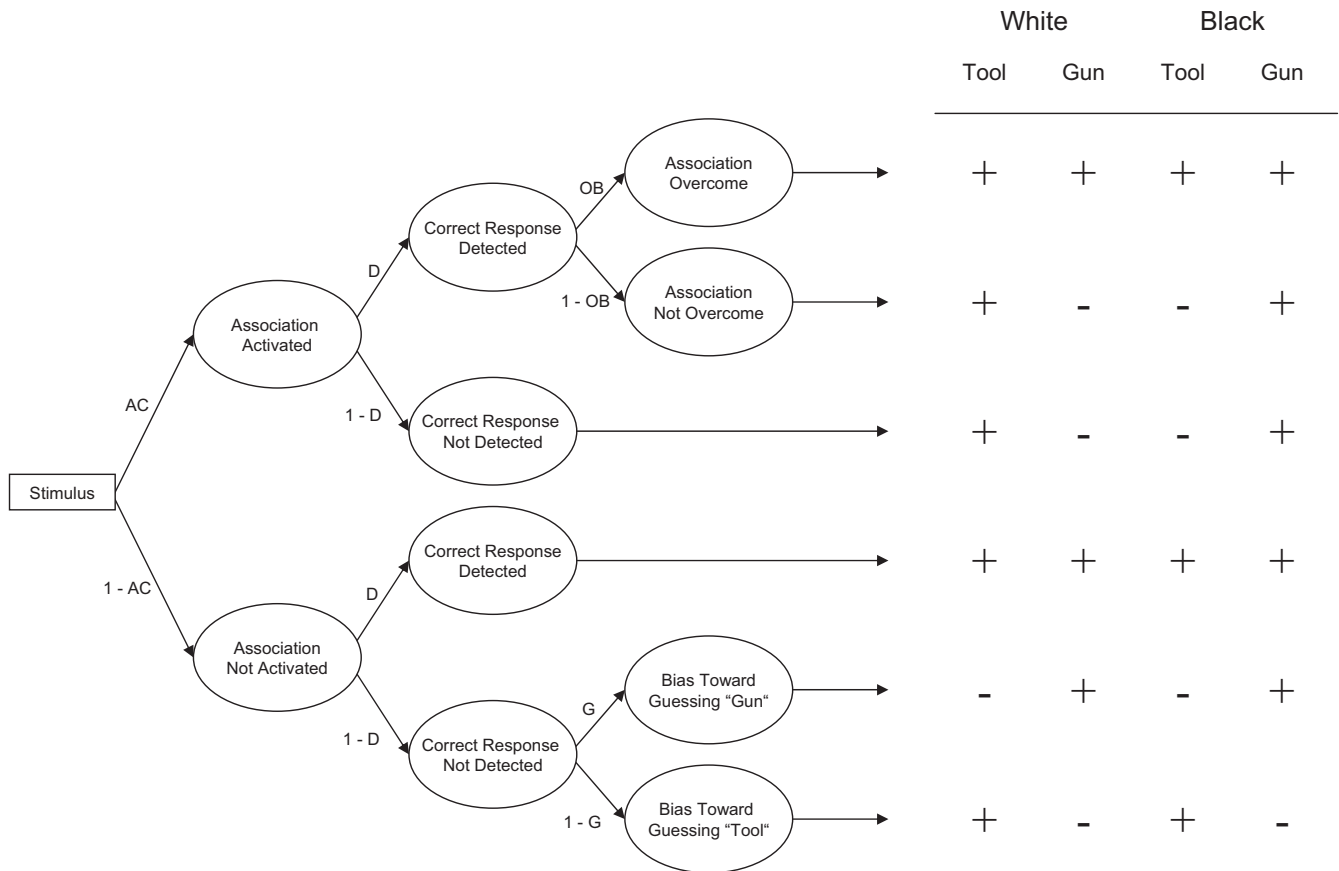


Figure 1. The Quadruple process model (Quad model), as applied to Payne's (2001) weapon identification task. Each path represents the likelihood of a correct response. Parameters with lines leading to them are conditional on all preceding parameters. The table on the right of the figure depicts correct (+) and incorrect (-) responses as a function of face prime (White vs. Black) and target stimulus (tool vs. gun). AC = activation; D = detection; G = bias toward guessing "gun"; OB = overcoming bias; 1 - AC = lack of association; 1 - D = failure of detection; 1 - G = bias toward guessing "tool"; 1 - OB = failure to overcome bias.

As an example of how the four processes in the Quad model interact to direct behavior, consider the shooter bias. As illustrated in Figure 1, different combinations of processes produce different behavioral outcomes, depicted by arrows moving from left to right. Initial presentation of a Black or White face may (AC) or may not (1 - AC) activate stereotypical associations. Participants also may (D) or may not (1 - D) be able to accurately detect that the target object is a gun or tool (D). If stereotypical associations are activated, then a key determinant of the behavioral outcome is whether the presented target object is compatible (Black-gun; White-tool) or incompatible (Black-tool; White-gun) with these associations. If the associations are activated and the stimulus object is compatible with these associations, then a correct response will be given, regardless of whether the correct answer can be detected. In this case, the associations alone are sufficient to provide a correct response, even if detection fails. However, if the associations are activated and the target object is incompatible with these associations, then whether a correct response is provided depends on the success of overcoming the associations (OB). If the associations are overcome (OB), then the correct response (as determined by the detection process) will be provided. If the

associations are not overcome (1 - OB), then they will drive behavior, and an incorrect response will result. Finally, if associations are activated (AC) but the correct response cannot be determined (1 - D), then the associations will determine behavior, leading to correct responses on compatible trials and incorrect responses on incompatible trials.

If the initial stimulus does not activate associations (1 - AC), then behavior is determined by the detection (D) and guessing (G) processes. In this case, if the correct answer can be detected (D), then a correct response will be given, regardless of the face prime and target object. However, if there is no activation (1 - AC) and detection fails (1 - D), then guessing determines the response. If respondents show a bias toward guessing "gun" (G), then they will provide correct responses on trials in which the target object is a gun and incorrect responses on trials in which the target object is a tool, regardless of whether the prime face is Black or White. In contrast, if there is a bias toward guessing "tool" (1 - G), then correct responses will be provided on trials in which the target object is a tool and incorrect responses will be provided on trials in which the target object is a gun, regardless of whether the prime face is Black or White.

### Analyzing Data With the Quad Model

Conrey et al. (2005) provided an elaborate description of the nature of multinomial modeling (see also Batchelder & Riefer, 1999; Klauer & Wegener, 1998; Riefer & Batchelder, 1988) and detailed instructions on how to analyze data with the Quad model. Thus, we refrain from an extensive discussion of these points. However, to facilitate a better understanding of the mathematical implementation of the Quad model, we briefly describe the basic logic of data analyses with our model.

Like other multinomial models, the Quad model generates estimates of its parameters from the proportions of correct and incorrect responses provided on different types of trials. The processing tree presented in Figure 1 illustrates how the model predicts correct and incorrect responses on compatible and incompatible trials as a function of the operations of four different processes. For example, there are three different ways to arrive at an incorrect response to a Black person with a tool (i.e., to respond "gun" instead of "tool"). First, an incorrect response may arise if stereotypical associations are activated (AC), detection succeeds (D) and overcoming bias fails ( $1 - OB$ ). Second, an error also would be the outcome if stereotypical associations are activated (AC) and detection fails ( $1 - D$ ). Finally, an error would be produced if stereotypical associations are not activated ( $1 - AC$ ) detection fails ( $1 - D$ ), and a bias toward guessing "gun" (G) produces an incorrect response.

Each of these three combinations of processes represents a set of conditional probabilities for which an incorrect response is produced. For instance, the likelihood that stereotypical associations are activated (AC), detection succeeds (D), and overcoming bias fails ( $1 - OB$ ) is represented by the product term  $AC \times D \times (1 - OB)$ . Thus, the overall likelihood of producing an incorrect response to a Black face with a tool is the sum of the three conditional probabilities:  $[AC \times D \times (1 - OB)] + [AC \times (1 - D)] + [(1 - AC) \times (1 - D) \times G]$ . The respective equations for each item category (e.g., a Black person with a gun; a White person with a tool) are then used to predict the observed proportion of errors in a given data set. Based on the full set of all equations, the Quad model estimates specific values for AC, D, OB, and G by means of maximum likelihood statistics. The predicted error rates implied by these estimates are then compared with actual error rates in the data set, which provide a chi-square fit-estimate for the accuracy of the model in predicting the observed data. More precisely, the four parameter values are changed through maximum likelihood estimation until they produce a minimum possible value of the chi-square. If the chi square test of model-fit is not significant, then the model is said to provide an accurate description of the data. However, if the chi-square test is significant, the predictions of the model significantly deviate from the observed data. The estimated parameter values resulting from this process are interpreted as relative levels of the four processes.

### VALIDATION AND APPLICATION OF THE QUAD MODEL

To date, the Quad model has been applied to and has been shown to accurately predict behavior on a variety of priming tasks, including

semantic priming tasks (Gawronski & Bodenhausen, 2005; Gonsalkorale, Sherman, Allen, Amodio, & Bartholow, 2007), evaluative priming tasks (Allen, Sherman, & Gonsalkorale, 2008), and the weapon identification task (Conrey et al., 2005; Gonsalkorale, Sherman, Allen, et al., 2007; Payne, 2001). The model also has been applied extensively to the IAT (Conrey et al., 2005; Gonsalkorale, Sherman, Allen, et al., 2007; Greenwald et al., 1998), the most widely used measure of automatic associations, and the Go/No-Go Association Task (GNAT; Gonsalkorale, von Hippel, & Sherman, 2007; Nosek & Banaji, 2001). Evaluative priming tasks are proposed to measure the extent to which positive and negative associations are automatically activated upon exposure to some stimulus. Responses on these tasks often are referred to as implicit attitudes. In contrast, semantic priming tasks and the weapon identification task are proposed to measure the conceptual knowledge that is automatically activated by some stimulus. Responses on these tasks often are referred to as implicit beliefs or implicit stereotypes. The IAT and GNAT can be used to measure either implicit attitudes or implicit beliefs. Though most of these measures have their roots in social and cognitive psychology, they have been adopted in almost all areas of psychology, including personality psychology (e.g., Asendorpf, Banse, & Mücke, 2002), developmental psychology (e.g., Rutland, Cameron, Milne, & McGeorge, 2005), cross-cultural psychology (e.g., Kim, Sarason, & Sarason, 2006), health psychology (e.g., Sherman, Rose, Koch, Presson, & Chassin, 2003), clinical psychology (e.g., Teachman, Gregg, & Woody, 2001), consumer psychology (e.g., Maison, Greenwald, & Bruin, 2004), forensic psychology (e.g., Gray, McCulloch, Smith, Morris, & Snowden, 2003), and neuropsychology (e.g., Beer et al., 2007; Phelps et al., 2000).

In each of these domains, these measures have been used to assess the prevalence and the effects of mental associations that are assumed to be activated automatically and without intention and whose expression cannot be altered or inhibited by controlled processes (e.g., Bargh, 1999; Devine, 1989; Fazio et al., 1995; Greenwald et al., 1998). Thus, self-regulatory processes have been seen as largely irrelevant to understanding responses on such measures. In contrast, we propose that both automatic and controlled processes underlie behavior on these tasks and that these processes can be independently measured with 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. For the adult to perform the task accurately, the automatic habit of reading the word must be overcome on incompatible trials (e.g., the word *blue* written in red ink). In contrast, the child has no automatic reading habit to overcome. The same logic applies to many evaluative and semantic priming tasks, to the IAT, and to the GNAT, which have the same compatibility structure as the Stroop task. As such, identical responses of two individuals on these measures may reflect moderately biased associations in the one case, and strong associations that are successfully overcome in the other. The Quad model can be used to disentangle these and other processes.

It is important however, to note that the implications of our research extend well beyond automatic activation and regulation of mental constructs during task performance. Because they create self-



regulation needs that mirror those encountered in everyday life, exploring the processes required to successfully perform these tasks can enhance understanding of how people achieve self-regulation over automatic associations and impulses in much broader contexts. For example, the extent to which a person has negative associations toward Muslims automatically activated or is able to overcome those associations during the course of completing a Muslim-related IAT might predict how he or she will get along with a Muslim person during on actual interaction. We describe just such a finding below. More generally, processes enacted during these tasks may shed light on how people resolve conflicts between immediate (e.g., escape the spider, smoke a cigarette, eat a donut) and longer term (e.g., overcome spider phobia, quit smoking, lose weight) goals. We further address the relationship of the Quad model to models of self-regulation and goal pursuit below.

### Validation of the Quad Model

The viability of the Quad model depends on four critical elements: model fit (i.e., Does the model adequately approximate behavioral data?), stochastic validity of the parameters (i.e., Can the model's parameters be influenced independently?), construct validity of the parameters (i.e., Do the parameters signify the processes claimed by the model?), and predictive validity of the parameters (i.e., Do the parameters predict meaningful behaviors?). The Quad model has succeeded on all fronts. As described above, the model has shown its ability to accurately predict performance on a variety of priming tasks, IATs, and the GNAT, demonstrating good model fit for these tasks (Allen et al., 2008; Conrey et al., 2005; Gonsalkorale, Sherman, Allen, et al., 2007; Gonsalkorale, von Hippel, Sherman, 2007).

#### *Stochastic Validity*

The stochastic validity of the model has been established in a number of ways (Conrey et al., 2005; Gonsalkorale, Sherman, Allen, et al., 2007). For example, implementing a response deadline in an IAT designed to assess implicit attitudes about flowers and insects reduced detection (D) and overcoming bias (OB), but left association activation (AC) and guessing (G) unaffected. Manipulating the base rate of left-hand responses versus right-hand responses in the same task affected guessing (G), but did not affect the other three parameters (AC, D, OB). The expectation that one's performance on the weapon identification task would be observed by others decreased the participants' ability to accurately detect the stimuli (D) but increased the success at overcoming bias (OB). These results indicate that the four parameters of the Quad model can vary independently, providing clear evidence for the stochastic validity of the model. Other examples of parameter independence are described below.

#### *Construct Validity*

The construct validity of the model parameters also has been established by a number of findings (Conrey et al., 2005; Gonsalkorale, Sherman, Allen, et al., 2007). The fact that detection (D) and overcoming bias (OB) were reduced by a response deadline supports the claim that the two parameters reflect controlled pro-

cesses that require cognitive capacity. In contrast, the finding that activation (AC) and guessing (G) were unaffected by the response deadline is consistent with their depiction as relatively automatic processes that do not require significant cognitive capacity. The validity of OB as a measure of self-regulation was further established by demonstrations that it is impaired by alcohol consumption and that it decreases with age (Gonsalkorale, Sherman, Allen, et al., 2007). Extensive research has shown that both alcohol use (e.g., Easdon & Vogel-Sprott, 2000) and aging (e.g., Hasher & Zacks, 1988) are associated with impairments in self-regulation. The fact that altering the base rate of left-hand and right-hand responses influenced G corroborates the portrayal of that parameter as a general response bias.

Finally, a neuroimaging study of performance on an evaluative IAT (Beer et al., 2007) showed that AC was correlated with activity in the amygdala and insula, which are involved in emotional processing and arousal (Britton, Taylor, Sudheimer, & Liberzon, 2006; Murphy, Nimmo-Smith, & Lawrence, 2003; Phan, Wager, Taylor, & Liberzon, 2002; Phelps et al., 2000). This finding is consistent with the depiction of AC as measuring evaluative associations in such a task. 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, Nystrom, Fissell, Carter, & Cohen), activity in the DLPFC has been linked to inhibitory control over prepotent responses (e.g., Chee, Sriram, Soon, & Lee, 2000; Taylor, Kornblum, Lauber, Minoshima, & Koeppel, 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 that feeds into efforts to overcome inappropriate automatic influences.<sup>5</sup> Altogether, there are now considerable behavioral and neuroscientific data indicating that the Quad model's parameters reflect the processes that they are intended to assess. Further examples are described below.

#### *Predictive Validity*

Two studies provide evidence for the predictive validity of the parameters. First, estimates of individual participants' association activation (AC) parameters derived from an evaluative IAT were positively correlated with association-related reaction time impairment in the same task (Conrey et al., 2005). Thus, the higher the AC, the greater the association-based impairment in performance. At the same time, estimates of participants' overcoming bias (OB) parameters were negatively correlated with association-based reaction time impairment. Thus, the higher the OB, the better able were participants to avoid association-based impairments in per-

<sup>5</sup> 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.

formance. These findings also bolster the construct validities of the AC and the OB parameters.

Second, in another study (Gonsalkorale, von Hippel, & Sherman, 2007), non-Muslim, Caucasian participants interacted with an experimental confederate who appeared to be and who was described as Muslim. Following the interaction, the confederate rated how much he or she liked the participants, while the participants completed a GNAT measuring implicit bias toward Muslims. The confederate's ratings of how much he or she liked the participants were predicted by an interaction between the AC and the OB parameter estimates taken from the GNAT. Specifically, when participants had low AC estimates of negative associations with Muslims, level of OB was unrelated to how much participants were liked by the confederate. In contrast, participants with high AC estimates of negative associations with Muslims were liked to the extent that the participants had high OB parameter estimates. Thus, the ability to overcome automatic negative associations on the GNAT predicted the quality of the social interaction when those associations were strong.

In sum, the Quad model has shown an ability to accurately describe behavior on a variety of evaluative priming tasks, semantic priming tasks, IATs, and the GNAT. In addition, the stochastic and construct validities of the model's parameters have been supported by numerous findings. Finally, the predictive validity of the AC and the OB parameters has been demonstrated.

### Applications of the Quad Model

The Quad model has been applied to a number of empirical and theoretical problems surrounding the automatic activation of mental associations (and related behavioral impulses) and their regulation. One set of questions pertains to understanding the contextual variability and malleability of responses on implicit measures of attitudes and beliefs. When responses change as a result of situational manipulations, what accounts for it? According to the Quad model, such effects could be due to changes in the nature of the activated associations, changes in respondents' ability to determine appropriate behavior, changes in respondents' ability to overcome automatic associations when necessary, changes in response biases, or some combination of these processes. An especially important practical concern is to understand exactly how various treatments and interventions designed to reduce associative biases produce effects. Another set of questions surrounds the meaning of individual and group differences in implicit attitudes and beliefs, including those associated with aging. To what extent do these differences reflect variation in underlying associations, the ability to determine appropriate behavior, the ability to overcome associations, response biases, or some combination of these processes? Finally, the model also has been applied toward understanding the underlying automatic and controlled processes that predict important extra-task behaviors in broad domain-relevant contexts. The following sections present applications of the Quad model to these issues, in turn: (a) accounting for the malleability of measures of automatic association, (b) accounting for variability in measures of automatic association, and (c) accounting for extra-task behaviors in domain-relevant contexts.

Distinguishing among processing accounts is important for a variety of theoretical and practical reasons, detailed below. Most

basically, clarifying these questions is critical in gaining a more complete understanding of what these measures assess and how they should be conceptualized. For example, the common interpretation that responses reflect only the automatic activation of associations may be an underestimation of not only the extent of controlled processing but also the extent of automatic activation, because a strong ability to overcome automatic bias may mask the true extent of that bias.

### *Malleability of Measures of Automatic Association*

People may be unaware of their attitudes or unwilling to report them truthfully. The "willing and able" issues are two of the most difficult problems for research on attitudes. The advent of implicit measures has offered promising new ways to avoid these obstacles by measuring attitudes and beliefs without directly requesting that respondents report those attitudes and beliefs. In many cases, people are unaware that their attitudes are being measured with such tasks. Many proponents of these measures further argue that even if made aware of the nature of the task, people are unable to control their responses. Thus, these measures are seen as reflecting the unintended, automatic activation of stored associations, whose expression, largely, cannot be altered or inhibited (e.g., Bargh, 1999; Devine, 1989; Fazio et al., 1995; Greenwald et al., 1998; Kim, 2003). Because responses on these measures are seen as impervious to short-term manipulation, they also are believed to reveal deep-seated, true attitudes that are consistent across situations and time (for a critical discussion, see Gawronski, LeBel, & Peters, 2007).

However, though they originally were assumed to be highly stable and resistant to change, considerable research now indicates that responses on these measures are highly context dependent (for reviews, see Blair, 2002; Gawronski & Bodenhausen, 2006; Gonsalkorale, Sherman, Allen, et al., 2007). For example, priming and IAT results are sensitive to subtle contextual features of the stimuli (e.g., Barden, Maddux, Petty, & Brewer, 2004; Wittenbrink, Judd, & Park, 2001), temporary changes in the accessibility of different features of the attitude object (e.g., Dasgupta & Greenwald, 2001; Govan & Williams, 2004; Mitchell, Nosek, & Banaji, 2003), variations in the context in which the measure is administered (e.g., Lowery, Hardin, & Sinclair, 2001; Sechrist & Stangor, 2001; Sinclair, Lowery, Hardin, & Colangelo, 2005; Richeson & Ambady, 2003), fluctuations in respondents' physical and motivational states (e.g., Ferguson & Bargh, 2004; Seibt, Häfner, & Deutsch, 2007; Sherman, Rose, et al., 2003), and many other factors.

Several researchers attributed these effects to contextual variability in the associations activated by a given stimulus or category (Blair, 2002; Ferguson & Bargh, 2007; Gawronski & Bodenhausen, 2006). Yet, other researchers have explained these findings by positing automatic inhibition of associative knowledge (e.g., Glaser & Knowles, 2008; Maddux, Barden, Brewer, & Petty, 2005; Moskowitz, Gollwitzer, Wasel, & Schaal, 1999). One important commonality among these explanations is that they all posit that malleability effects are due, in one way or another, to variation in the associations automatically activated during task performance. Such explanations would seem to be a logical necessity, given the predominant view that responses on these mea-

tures reflect only the automatic activation of stored associations and cannot be altered or inhibited by nonassociative responses.

In contrast to these accounts, in the Quad model, it is proposed that contextual variability could also be due to changes in controlled processes, namely, the ability to determine appropriate behavior or overcome automatic associations when necessary. From this perspective, accounts of response malleability may have significantly overstated the ease of altering underlying associations and underestimated the role of controlled intentions and motivations. Of course, different interventions aimed at changing attitudes, beliefs, and behaviors may produce effects through different processes. The Quad model can help to specify these processes and to promote better understanding of which kinds of interventions are most likely to be successful under which circumstances. To examine these issues, we have applied the model to a number of cases in which implicit evaluative and conceptual biases were altered by experimental manipulations.

### *Changing Newly Formed Attitudes*

In some cases, changes in implicit evaluations appear to primarily reflect changes in the underlying associations that are activated automatically. One of the most basic demonstrations of such an effect concerns the situation in which attitudes are formed toward a novel object and subsequently altered as new information is acquired. As an example, we replicated a study by Rydell and McConnell (2006), in which participants learned about a target person named Bob (Gonsalkorale & Sherman, 2007). Bob was described with 100 positive behaviors, leading to the formation of positive implicit (as measured by an IAT) evaluations of Bob. Following the initial induction of the positive impression, participants read descriptions of 0, 20, or 100 negative behaviors performed by Bob, after which, evaluations were measured again. Implicit evaluations of Bob changed in response to the new information, with evaluations becoming increasingly negative as the number of Bob's negative behaviors increased. Application of the Quad model revealed that AC estimates of positive, automatic associations with Bob diminished as negative information accumulated. No other parameters were affected by the negative information. Thus, diminishing positive IAT scores were associated with diminishing positive associations with Bob.

### *Exemplar Accessibility*

Another case in which variations in implicit evaluations are associated with variations in underlying associations involves the differential accessibility of category exemplars. A number of studies have shown that implicit evaluations of social groups (e.g., Black and White people) are affected by the particular group members that are currently salient (e.g., Blair, Ma, & Lenton, 2001; Dasgupta and Greenwald, 2001; Govan & Williams, 2004; Mitchell et al., 2003). For example, Govan and Williams (2004) showed that anti-Black implicit bias on an IAT was significantly weaker when the stimuli were faces of positive Black and negative White exemplars than when the stimuli were faces of unknown Black and White targets. We replicated this study (Gonsalkorale, Sherman, Allen, et al., 2007) and applied the Quad model to the data.

Because the IAT is a measure of relative evaluations of two targets, estimating the independent evaluative associations for each target is impossible in the standard reaction time analysis (Nosek, Greenwald, & Banaji, 2005). An advantage of the Quad model is that it can produce separate estimates for different types of associations. As such, we created one AC parameter that measured the extent to which associations between *Black* and *unpleasant* were activated in performing the task, and another AC parameter that measured the extent to which associations between *White* and *pleasant* were activated in performing the task. Modeling results showed that the only parameters that differed between the conditions were the two AC parameters. Specifically, when the IAT presented positive Black and negative White exemplars, the activated Black associations were more positive and the activated White associations were more negative than when the IAT included unknown Black and White exemplars. Thus, implicit evaluations of social groups depend on the particular group exemplars that are currently salient and on the associations activated by those exemplars.

### *Training to Negate Biased Associations*

Other interventions that change implicit evaluations influence both automatic and controlled processes. Previous research showed that participants who were trained to negate stereotypes showed significantly less implicit stereotyping on a variant of the Stroop task than did participants who were trained to maintain stereotypes or who received no training (Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000; but see Gawronski, Deutsch, Mbirkou, Seibt, & Strack, 2008). We conducted a similar study, in which participants completed a task that trained them to negate or maintain anti-Black and pro-White evaluative associations before performing an evaluative IAT (Gonsalkorale, Sherman, Allen, et al., 2007). Participants in the negate associations condition were instructed to press a *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 a *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 (i.e., to press the *YES* key in response to Black and negative pairings or White and positive pairings and to press the *NO* key in response to Black and positive pairings or White and negative pairings). After completing 480 trials of such training, participants completed an evaluative IAT. IAT results showed that the training was effective in reducing bias: Those trained to negate associations showed less IAT bias than those trained to maintain associations. Analysis with 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). According to Monteith and her colleagues (Monteith, Ashburn-Nardo, Voils, & Czopp, 2002), behavioral monitoring is an essential skill in responding without bias because successfully discriminating appropriate actions from inappropriate actions is a necessary precondition for regulating behavior. Our modeling results show that one of the benefits of antibias training is an increased ability to monitor appropriate behavior.

### *The Impact of Public Versus Private Contexts on Implicit Bias*

Some situational contexts influence multiple automatic and controlled components of implicit bias. A particularly useful feature of the Quad model is the ability to shed new light on already published data. In so doing, the model often can provide a more nuanced and detailed description of previous research findings. One example is a reanalysis of a study by Lambert et al. (2003) on the impact of anticipated public contexts on implicit stereotyping (see Conrey et al., 2005). Lambert, Cronen, Chasteen, and Lickel (1996) showed that an anticipated public context, compared with a private condition, ironically increased (rather than decreased) the expression of racial bias. This finding seems quite surprising, given that making people accountable to others for exhibiting socially undesirable biases may be expected to diminish the extent of such biases (e.g., Lerner & Tetlock, 1999).

Lambert et al. (2003) suggested two possible explanations of these findings. The first is a habit-strengthening or drive-based explanation. According to this explanation, public contexts increase arousal, and this arousal leads to an increase in the dominant response (Hull, 1943; Zajonc, 1965). Given that stereotypic associations reflect a particular kind of dominant response, public contexts may lead to a higher activation level of stereotypical associations than do private contexts, leading to greater stereotypical bias. The second possible explanation offered by Lambert et al. (2003) is an impairment of control account. According to this explanation, public contexts divide participants' attention (Baron, 1986), thereby reducing cognitive capacity. Thus, because determining correct responses on the task is cognitively effortful, public contexts may decrease this ability, which in turn should lead to greater influence of stereotypic associations under public conditions than under private conditions. On the basis of an application of Jacoby's (1991) process dissociation procedure, Lambert et al. (2003) concluded that public accountability led to an impairment of control but not to an increase in automatic activation of the stereotype.

In contrast to this conclusion, our application of the Quad model to these data indicated that the influence of anticipated public scrutiny is more complex (see Conrey et al., 2005, Experiment 5). First, our analysis showed that detection (D) was lower under public conditions than under private conditions. This finding is consistent with distraction-conflict models of social facilitation (e.g., Baron, 1986), suggesting that public contexts create attentional conflicts, thereby reducing the ability to accurately detect the presented stimuli. Second, we found that overcoming bias (OB) was higher under public conditions than under private conditions. This result is consistent with the expectation that being accountable to others increases people's motivation to overcome their stereotypical associations. These findings indicate that public accountability does not impair all aspects of self-control. Though it does reduce the ability to determine correct responses, it also increases people's ability to overcome unwanted bias. More broadly, this result is important because it shows the value of separating different types of control that may be influenced in opposite ways by the same context. Finally, we found that the activation of stereotypical associations (AC) was higher under public conditions than under private conditions. This finding is

consistent with the habit-strengthening account proposed by Lambert et al. (2003), suggesting that public contexts increase arousal, which in turn leads to an increase in the dominant response (Hull, 1943; Zajonc, 1965): in this case, the activation of stereotypic associations. This finding was obscured by Lambert et al.'s (2003) measure of automatic stereotype activation, which does not separate the extent of activation from the extent to which associations are successfully overcome (more on this below).

### *The Effects of Alcohol on Implicit Bias*

To this point, all the research we have described shows that variability in implicit attitudes and beliefs is associated, at least in part, with variability in the nature of the associations automatically activated during responding. However, in some cases, variations in implicit bias appear to have nothing to do with the underlying associations but rather reflect only variations in controlled processes.

Research has indicated that alcohol impairs cognitive and motor performance by reducing the ability to regulate prepotent responses. For example, intoxicated individuals are less able than others 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, some participants were given alcohol prior to completing a priming measure of implicit racial stereotyping. Results indicated that a high dose of alcohol, which was compared with a placebo, increased stereotypic responding on the task. Applying the Quad model to these data, we found that overcoming bias was the only parameter that differed across alcohol consumption conditions. This finding suggests that alcohol intoxication interferes with people's ability to regulate automatically activated associations but does not alter the nature of those associations.

### *Variability in Measures of Automatic Association*

Just as malleability of implicit attitudes and beliefs may be based on a variety of different processes, so too may individual and group differences. The example of an adult and child producing identical Stroop task performance for different reasons illustrates this possibility and the difficulty of interpreting behavioral data. Numerous studies demonstrate individual differences on measures of implicit attitudes and beliefs. For example, individuals who are high in chronic egalitarian goals (Moskowitz et al., 1999; Moskowitz, Salomon, & Taylor, 2000) or who are internally but not externally motivated to respond without prejudice (Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002) exhibit less implicit bias and stereotyping than do other individuals. Implicit attitudes also are a function of the perceiver's group membership. The robust finding of ingroup bias on explicit measures (Hewstone, Rubin, & Willis, 2002) is often reflected on implicit measures, as well (e.g., Ashburn-Nardo, Voils, & Monteith, 2001; Greenwald et al., 1998; Perdue, Dovidio, Gurtman, & Tyler, 1990; Rudman, Greenwald, Mellott, & Schwartz, 1999). Finally, research has identified age as another important source of individual differences, with older



adults exhibiting greater implicit racial bias and gender stereotyping than do younger adults (Nosek, Banaji, & Greenwald, 2002). Variation in implicit bias may thus arise from individual differences in motivations, goals, group memberships, and age.

The core premise of much of this research is that responses on measures of implicit attitudes and beliefs reflect only the associations that are automatically activated and are not subject to intention or control. Given this understanding, any variations in task performance observed among respondents must, by definition, reflect differences in activated associations. In contrast, according to the Quad model, different performance may reflect variation in underlying associations, the ability to determine appropriate behavior, the ability to overcome associations, response biases, or some combination of these processes.

Better specifying the sources of variability on these tasks may help in better identifying means of changing relevant attitudes, beliefs, and behaviors. If biased behavior stems from the nature of the associations activated then interventions that directly influence those associations may be most effective for changing the behavior. In contrast, if the biased behavior stems from deficits of self-regulation then interventions that improve self-regulation may be most effective. Thus, matching change strategies with the appropriate situations is an important tool for reducing undesirable, impulsive behaviors. The Quad model can help to identify the underlying bases of automatic biases, as well as the specific mechanisms through which different interventions work (as described in the previous section on malleability).

### *Bases of Ingroup and Outgroup Implicit Bias*

In one study (Gonsalkorale, Sherman, Allen, et al., 2007), we examined the factors underlying the common finding that Black people show less positivity toward Whites and less negativity toward Blacks on implicit measures than do White people (e.g., Nosek et al., 2002). A possible basis for this finding lies in the cultural environments of Black and White Americans (Entman & Rojecki, 2001). Given that subcultures within a society tend to emphasize positive ingroup exemplars (e.g., Simonton, 1998), Black people should generally encounter fewer positive White and fewer negative Black exemplars than do White people. This differential exposure to exemplars should be reflected in the associations that are automatically activated among Black and White participants, just as our experimental manipulation of exemplar positivity (described above) affected the associations that are activated. Thus, we predicted that Black participants would show lower levels of pro-White and anti-Black association activation than would White participants. We did not expect Black participants' lower levels of implicit bias to be related to detection, overcoming bias, or guessing. Results supported this hypothesis, with White participants showing stronger automatic activation of both Black-negative and White-positive associations but showing no other parameter differences.

### *Individual Differences in Motivation to Respond Without Prejudice*

In other cases, interpersonal variation in implicit bias is related to both automatic and controlled processes. Considerable research

(Amodio, Devine, and Harmon-Jones, 2008; Amodio, Harmon-Jones, & Devine, 2003; Devine et al., 2002) has shown that individuals who are internally but not externally motivated (high internal motivation to respond with prejudice [IMS]/low external motivation to respond without prejudice [EMS]) to behave in nonprejudiced ways demonstrate less bias on measures of implicit bias than do individuals who are motivated by both internal and external reasons (high IMS/high EMS participants) or who lack internal motivation (low IMS). However, relatively little is known about how they achieve nonbiased responding on such measures. To examine this issue, we applied the Quad model to the data from the Amodio et al. (2008) weapons identification task. We also conducted a new study in which we looked at the effects of internal and external motivations on performance of an evaluative IAT. Both studies showed that high IMS/low EMS participants showed less implicit bias than did other participants. Quad model analyses of the data showed that compared with the other participants, high IMS/low EMS participants showed less activation of biased associations (AC) in performing both tasks. These participants also were more able to detect appropriate and inappropriate responses (D) on both tasks. There was no evidence of differences in overcoming bias (OB) as a function of different motivations in either study. These findings are consistent with those described above, in which directed training to negate biases reduced AC and enhanced D. One implication is that high IMS/low EMS individuals may be training themselves to behave in a nonbiased fashion, an implication that is consistent with theoretical accounts of IMS and EMS (Amodio, 2008; Amodio et al., 2003; Devine et al., 2002; Monteith et al., 2002). Whatever the origins, high IMS/low EMS individuals' ability to regulate implicit bias is associated with less biased automatic associations and enhanced ability to detect when regulation is required (i.e., when there are conflicting responses and a danger of responding inappropriately).

### *Aging and Implicit Bias*

Other individual differences in implicit bias appear to have nothing to do with activated associations and are, seemingly, based entirely on variations in controlled processes. Recent research has revealed a developmental trend, showing a positive correlation between age and implicit racial bias among White Americans (e.g., Nosek et al., 2002). This finding is often interpreted as evidence that older people's racial associations are more biased than are those of younger adults, reflecting generational changes in societal attitudes.

However, an alternative explanation for age differences in prejudice is that deficits in self-regulatory ability alter the attitudinal expression of older adults (von Hippel, Silver, & Lynch, 2000). Given that the ability to inhibit automatically activated stereotypes enables people to behave nonprejudicially (Bartholow et al., 2006; Devine, 1989; Moskowitz et al., 1999) and that inhibitory functioning declines with age (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.

We conducted a study to examine whether inhibitory processes can account for age differences in implicit racial bias. 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 from approximately 16,000 respondents as a function of participant age, which ranged from 11 years to 94 years. 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 race bias on the IAT, the older adults demonstrated less biased automatic associations (AC) than did the younger adults. The older participants also showed a greater likelihood of detecting (D) the correct response and a stronger propensity to make positive guesses (G) on the task. However, as predicted, overcoming bias (OB) decreased with age. It appears that despite weaker activation of associations, greater detection of correct responses, and positive guessing bias, the older adults exhibited stronger implicit bias behaviorally because they were less able to inhibit their activated associations. These findings indicate that age differences in implicit racial bias are due to age-related losses in regulatory functions.

### *Quad Model Insights on Extra-Task Behavior*

We have argued that application of the Quad model to measures of implicit attitudes and beliefs holds promise for increasing understanding of broader goal-directed behavior. We believe that the automatic and controlled processes that interact to direct performance on these immediate response tasks is likely to predict success at resolving impulse regulation conflicts within the same domain between immediate, low-level, narrow goals (e.g., escape the spider; smoke a cigarette; eat a donut) and longer term, high-level, global goals (overcome spider phobia, quit smoking, lose weight). Thus far, we have produced evidence of such a relationship in the domain of interpersonal interactions.

Specifically Gonsalkorale, von Hippel, and Sherman's (2007) used the Quad model to understand interpersonal interactions between Caucasian participants and a Muslim confederate. In that study, the extent to which the confederate liked the participants was related to the participants' ability to overcome automatic anti-Muslim associations (OB) when performing a GNAT. However, the participants' ability to overcome bias mattered only when she or he had high levels of anti-Muslim associations in the first place (AC). When automatic associations were weak, OB was not critical to a smooth interaction.

This study shows that the Quad model parameters estimated from performance on measures of implicit attitudes and beliefs can predict domain-relevant behavior in a broader context. In ongoing research, we are also examining this relationship in the context of cigarette smoking. Research has shown that smokers' implicit attitudes about cigarettes are less negative than are those of ex-smokers and nonsmokers (Sherman, Rose, 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 nonsmokers. Alternatively, it may be that smokers are less able to determine appropriate smoking-related behaviors than are nonsmokers, or that smokers 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-regulation 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-regulation issues in many different domains of judgment and behavior.

### Summary

The Quad model has been extensively validated and has been applied to a number of important empirical and theoretical questions pertaining to automatic processes and their regulation. The model has been shown to accurately predict behavior on a variety of priming tasks and implicit association tasks (IAT, GNAT) measuring both automatic evaluations and beliefs. Findings support the stochastic, construct, and predictive validities of the model's parameters. The model has been used to specify the processes underlying change and malleability of implicit attitudes and beliefs. Counter to widely held views, modeling analyses showed that these effects are not related solely to changes in activated associations. Though in some cases, these effects are related solely to variations in activated associations (changing novel attitudes; altering exemplar accessibility), in other cases, the controlled process of detection (training; public vs. private contexts) is also related. In still other cases (the effects of alcohol), implicit attitude malleability may occur in the absence of changes in activated associations and is related only to variations in the controlled process of overcoming bias. The model also has been applied toward understanding group, individual, and developmental differences in implicit biases. Here, too, the model showed that these differences are sometimes related to variability in activated associations (e.g., Black participants vs. White participants), sometimes related to variability in both activated associations and controlled processes (IMS and EMS), and sometimes related solely to controlled processes (aging effects). Finally, parameter estimates of the model derived from implicit measures have been used to predict domain-relevant behavior in the broader context of interpersonal behavior.

### COMPARISONS WITH OTHER MODELS

In the remainder of this article, we compare the Quad model with current models of self-regulation and with prominent dual-process models of automaticity and control. These comparisons are important for situating the model in the broader literatures on self-regulation and on automaticity and control and for highlighting the theoretical and empirical advances afforded by the Quad

model. In doing so, these comparisons highlight the conditions and the purposes for which the Quad model may be of use.

### Comparison With Models of Self-Regulation

Many questions pertaining to self-regulation are concerned with instances in which self-control is aimed at resolving conflicts between lower level and higher order goals (e.g., satisfying a sugar craving vs. staying on a diet; Fujita, Trope, Liberman, & Levin-Sagi, 2006; Loewenstein, 1996; Metcalfe & Mischel, 1999; Muraven & Baumeister, 2000; Trope & Fishbach, 2000). The Quad model proposes that self-regulation processes are also crucial whenever a situation is characterized by competition between automatic associations (and their related impulses) and responses that promote goal attainment. These automatic associations need not be goal-related. Thus, a police officer does not need to have a goal to associate Black people with aggression; this association simply interferes with his ability to take appropriate action toward an unarmed Black suspect. When driving in Great Britain, an American has no goal to drive on the right side of the road; it is simply an automatic habit that may hinder safe driving. Finally, when completing the Stroop task, a person has no goal to read the word, though the reading habit interferes with naming the color of the ink.

The implicit measures to which the Quad model has been applied similarly place automatic associations into conflict with a detected appropriate 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 on these tasks, this conflict does not arise from competing goals. Though other models similarly assume a role for self-regulation in conflicts between impulses and opposing behaviors (e.g., Fujita et al., 2006; Metcalfe & Mischel, 1999; Muraven & Baumeister, 2000), the Quad model is the first to apply this analysis within such compatibility tasks.

As a general model of impulse regulation, the Quad model applies to any context in which automatic, impulsive, or prepotent response tendencies and intended, more controlled, contextually appropriate responses may conflict. Thus, as discussed above, theoretically, the Quad model also applies to self-regulation contexts with competing low-level and high-level goals (e.g., phobic behavior, smoking, dieting, intergroup behavior, etc.). An important question in the ongoing development of the model will be the extent to which parameter estimates derived from implicit task performance predict behavior in broader self-regulatory contexts. We have shown initial evidence that such parameter estimates do, in fact, offer important insights into the self-regulation of competing goals.

The Quad model's ability to quantify the relative contributions of the four different processes also provides an important extension to prominent narrative models of impulse regulation (e.g., Fujita et al., 2006; Loewenstein, 1996; Metcalfe & Mischel, 1999; Moskowitz et al., 1999; Muraven & Baumeister, 2000; Strack & Deutsch, 2004; Trope & Fishbach, 2000). Whereas the major contribution of these models lies in their predictions of how specific factors influence self-regulation, the primary contribution of the Quad model lies in its ability (a) to specify how different

processes interact in producing a particular behavioral outcome and (b) to quantify the relative contribution of these processes within a given task or context. Self-regulation models propose important roles for both impulse strength and control strength. However, these models do not provide any means to quantify the relative contributions of the two. Thus, without a quantification of the two factors, a specific behavioral outcome (e.g., not smoking) may reflect the presence of a strong impulse that is successfully overcome (e.g., success at controlling a strong urge to smoke) or the complete absence of any impulse (e.g., a weak urge to smoke). To the extent that self-regulation is effective, this may lead to both an underestimation of the extent of successful self-regulation and an underestimation of the full strength of the impulse. Hence, the Quad model not only provides more fine-grained analyses of the processes underlying impulse regulation by going beyond the standard dichotomy of automaticity versus control but also provides an important tool for assessing the relative contributions of the proposed processes to a particular behavioral outcome.

### Comparison With Dual-Process Models of Automaticity and Control

In this section, we compare the Quad model with dual-process models of automaticity and control. Currently, such models are the primary accounts of how automatic and controlled processes interact to direct behavior. Though they are not always explicitly presented as such, dual-process models almost always are relevant to questions of self-regulation in that they are concerned with delineating the circumstances under which judgment and behavior are driven by automatic, unintended processes versus sometimes competing, controlled, intended processes.

The most obvious difference between the Quad model and the dual-process models is that the Quad model specifies four qualitatively distinct processes rather than the basic automatic-controlled dichotomy. Including all four processes in a single model increases the breadth of application of the model. Many dual-process models are limited to explaining behavior within the specific content domains for which they were developed (e.g., judgment and decision making, attribution, stereotyping; for a discussion, see Smith & DeCoster, 2000). Because dual-process models specify one automatic process and one controlled process, those models are not able to generalize to domains in which other qualitatively distinct processes are important.

Though many specific dual-process models have been proposed to account for behavior in different domains, these models generally have relied on one of three broad approaches for describing the interactions between automatic processes and controlled processes: content dissociation, task dissociation, and process dissociation. We discuss these different approaches in turn. In addition, we discuss dual-process models that place automatic and controlled processes within specialized neurocognitive systems.

#### *Content Dissociation Approaches*

In many domain-specific dual-process models, the interaction of automaticity and control is tied to particular types of information in judgment tasks. For example, prominent dual-process models of judgment and decision making (Tversky & Kahneman, 1974),

persuasion (Chaiken, 1980; Petty & Cacioppo, 1981), impression formation (Brewer, 1988; Fiske & Neuberg, 1990), and dispositional attribution (Gilbert, Pelham, & Krull, 1988) are based on this approach. In these models, the use of one kind of information is thought to reflect a relatively automatic process (e.g., judgmental heuristics, peripheral cues, stereotypes, dispositional attributions), and the use of another kind of information is thought to reflect a more controlled process (e.g., statistical reasoning, argument strength, individuating information, situational attributions). Regulation of automatic processes is achieved via the use of alternative, more controlled sources of information.

One drawback of these models is that they tend to confound the influence of different pieces of information (content) with processing assumptions about automaticity and control. Though some kinds of information (e.g., heuristic cues, stereotypes) may often be accessed and applied more easily than other kinds of information (e.g., persuasive arguments, individuating information), it also is possible to reverse the situation (e.g., Erb et al., 2003; Krull & Dill, 1996; Kunda & Thagard, 1996; Trope & Gaunt, 2000). Which kind of information is more easily accessed and applied often depends on the task, the configuration of the information, the context, or the perceivers' goals. From this perspective, the equation of specific types of content with extrinsic process features (i.e., automatic vs. controlled) seems problematic.

These models also have difficulties in accounting for the joint contributions of automatic and controlled processes to behavior. Some suggest that automatic and controlled processing represent two ends of a continuum (e.g., Fiske, Lin, & Neuberg, 1999; Fiske & Neuberg, 1990). Here, it is possible for behavior to reflect the joint influence of automatic and controlled processes, but it is impossible to determine the contribution of each. Movement along the continuum may reflect increased or decreased automatic processing, increased or decreased control, or a combination thereof. Another problem for continuum models is that they logically require that as automatic processes are enhanced, controlled processes are diminished, and vice versa. However, given that automatic and controlled processing are frequently independent of one another (e.g., Jacoby et al., 1993) or even positively correlated (e.g., Jacoby, Begg, & Toth, 1997), this is a significant drawback. These considerations all point to the advantage of assessing the contributions of qualitatively different processes simultaneously and independently, as the Quad model does.

Other dual-process content models propose that automatic and controlled processes represent distinct alternatives that do not co-occur (e.g., Brewer, 1988; Fazio, 1990; Petty & Cacioppo, 1981). Obviously, these kinds of models are not well suited for examining the joint contributions of automatic and controlled processes to behavior. To solve this problem, more recent versions of these models typically dissociate content and process (Brewer & Feinstein, 1999; Fiske et al., 1999; Petty & Wegener, 1999), proposing that all kinds of information (heuristics, stereotypes, statistical information, individuating information) may be processed in either an automatic or controlled manner. Many of these models also incorporated the idea that automatic and controlled processes occur simultaneously and interact with one another (e.g., Brewer & Feinstein, 1999; Fazio & Towles-Schwen, 1999; Petty & Wegener, 1999). However, automatic and controlled processes are still tied to particular content in that within any given behavior,

each process is associated with a particular type of information. Thus, a heuristic may be used in either an automatic or a controlled fashion, depending on the context. However, the use of that heuristic never represents a combination of automatic and controlled processes. Joint influences of automaticity and control are represented solely by the simultaneous influence of two different kinds of information (e.g., one used automatically and one used in a controlled fashion). This may be contrasted with the Quad model's assumption that the use of any type of information reflects a combination of different processes, some of which may be largely automatic and some of which are more controlled, and that these processes can be estimated simultaneously from a single response, independently of particular content.

### *Task Dissociation Approaches*

In a second class of dual-process models, researchers seek to assess automatic and controlled processes by administering two separate measures, one aimed at tapping an automatic process and one aimed at tapping a controlled process. Whereas the automatic measure is thought to be relatively immune to the influence of intentional control, the controlled measure is affected by such processes. This approach has become increasingly prevalent in the past 20 years. For example, in memory research, automatic memory effects are frequently assessed with various implicit measures (e.g., word-fragment or stem completion tasks), whereas controlled memory is assessed with explicit measures (e.g., free recall or recognition tasks; for a review, see Richardson-Klavehn & Bjork, 1988). Similarly, much research on attitudes assesses automatic attitudinal responses with implicit measures (e.g., priming tasks, IATs) and controlled expressions of attitudes with explicit measures (e.g., standardized questionnaires, feeling thermometer ratings; e.g., Devine, 1989; Fazio et al., 1995; Greenwald et al., 1998; Wittenbrink, Judd, & Park, 1997). Different responses on the two types of measures have been taken as evidence for distinct automatic and controlled processes (e.g., Fazio et al., 1995), mental representations (e.g., Greenwald et al., 1998; Wilson, Lindsey, & Schooler, 2000), or processing systems (e.g., Rydell & McConnell, 2006; Strack & Deutch, 2004) that operate independently of one another.

Although this task dissociation approach has been responsible for many significant advances across fields of psychology, it has certain limitations. First, it confounds the processing style (automatic vs. controlled) with the particular measurement task (e.g., implicit vs. explicit). This confounding is problematic because the chosen tasks may differ in a number of ways beyond the extent to which they tap automatic versus controlled processes, representations, or systems. For example, dissociations between implicit measures and explicit measures of memory have been reinterpreted as dissociations between measures of perceptual processing and conceptual processing (e.g., Roediger, 1990). In a similar vein, dissociations between implicit measures and explicit measures of attitudes have been reinterpreted as being due to their lack of structural correspondence (e.g., Payne, Burkley, & Stokes, 2008; see also Gawronski et al., 2007).

The more general point is that no task is process pure. Any task that requires an observable response (e.g., a button press) cannot rely entirely on automatic processes. At the least, control is re-

quired to convert activated content into a button press. In addition, given their ubiquitous nature, no task is immune from all automatic processes. Thus, all tasks (and the behaviors they are meant to represent and measure) are influenced jointly by simultaneously occurring automatic and controlled processes. Attempts to isolate the influences of automatic and controlled processes within separate measures will necessarily result in a misspecification of each component, and such attempts cannot identify the joint contributions of the components.

The Quad model resolves these problems by examining the relative contributions of multiple, qualitatively different processes within the same task. Rather than equating responses on a given task (e.g., implicit vs. explicit tasks) with a particular process feature (i.e., automatic vs. controlled), the Quad model assumes that no task is process-pure. By quantifying the impact of different processes within a single task, the model resolves the unwarranted equation of task and processing mode implied by task dissociation models. The extensive Quad model analyses of implicit measures described above attest to the multicomponential nature of tasks that are often presumed to reflect only automatic processes.

### *Process Dissociation Approaches*

The problems implied by task dissociation approaches led Jacoby (1991; Lindsay & Jacoby, 1994) to develop process dissociation (PD) techniques for estimating the impact of different processes within a single task. As with the Quad model, this approach avoids both the content–process confounds of content dissociation approaches and the task–process confounds of task dissociation approaches. The PD models also permit the simultaneous estimation of different processes within a single task. Indeed, the development of the Quad model was strongly influenced by the logic and the procedures of the PD approach, and the Quad model thus owes a significant intellectual debt to those models. However, the Quad model also differs from the PD models in important ways.

Jacoby and his colleagues initially developed two complementary models of process dissociation. Both models rely on contrasting compatible trials (in which an automatic and a controlled process should lead to the same response) with incompatible trials (in which an automatic and a controlled process should lead to different responses). The primary difference between the models is whether automatic or controlled processes are assumed to be primary.

### *The Recollection–Accessibility–Bias Model*

One model, the recollection–accessibility–bias (RA) model is designed for tasks in which an automatic process is thought to operate only in the wake of failed control. This is the late type of bias that is related to the G parameter in the Quad model. As an example, this model is used to independently estimate the roles of recollection and feelings of familiarity in performing a recognition memory task (e.g., Jacoby, 1991). The model assumes that familiarity influences responses only when recollection has failed to provide a response. This model of process dissociation has been successfully applied across a number of domains (e.g., Ferreira et al., 2006; Jacoby, 1999; Payne, 2001; Payne, Lambert, & Jacoby, 2002; Sherman, Groom, Ehrenberg, & Klauer, 2003; Toth, Reinhold, & Jacoby, 1994).

One important difference between this model and the Quad model is that the RA model permits no role for processes that capture attention and influence behavior, even though the correct response can be determined. Thus, the RA model is not applicable to tasks such as the Stroop task, in which people can usually determine the correct response easily but in which a habit of reading the word can drive behavior away from that response (for more thorough discussions, see Conrey et al., 2005; Sherman, 2006b).

### *The Inhibition Deficit Model*

The other PD model (the inhibition-deficit or ID model) is designed for attention capture paradigms, such as the Stroop task, in which activated associations or impulsive response tendencies may capture behavior from the outset, regardless of whether control succeeds (e.g., Lindsay & Jacoby, 1994). In this case, control influences behavior only in the absence of an automatic bias.

One important difference between this model and the Quad model is that the ID cannot distinguish between cases in which an association or impulse is not activated at all and cases in which an association or impulse is activated but successfully overcome. More precisely, the structure of the ID model implies that if an association or impulse is activated, it will drive behavior; there is no possibility of successfully regulating an activated bias. Thus, the ID model cannot distinguish between the adult and the child who produce identical performance on the Stroop task for different reasons. Another difference between the ID model and the Quad model is that the ID model includes no role for late automatic processes, such as response biases. According to the model, if there is no initial automatic bias and control fails, the result will always be an incorrect response.<sup>6</sup>

### *Choosing a Model*

The Quad model and the PD models share many important features. In each model, it is assumed that different processes operate simultaneously and independently. In addition, it is agreed that different processes influence behavior in an interactive manner. Finally, the models concur that the way to quantify these components is through opposition tasks that vary the compatibility between two response tendencies. How, then, should one choose which model to use? There are a number of ways to answer this question.

*Theoretical considerations.* One important basis for choosing a model is theoretical considerations. If a researcher is interested in an automatic process that captures attention and influences behav-

<sup>6</sup> Recently, Jacoby and his colleagues (e.g., Jacoby, Bishara, et al., 2005) have proposed a new PD model that incorporates both early and late automatic processes. However, the model's controlled process does not distinguish between discrimination processes and suppression processes (as does the Quad model) and, as a result, the model cannot separate strength of activation from ability to overcome activation. Jacoby, McElree, and Trainham (1999) created a counter model for the Stroop task in which they did consider both activation and suppression processes. In this model, suppression was represented as a gating function for the suppression of the word-reading habit in the Stroop task. However, this gating function was implemented as an input component contributing to the extent of the word-reading habit rather than as an independent process that is measured in its own right.



ior regardless of whether control succeeds, then the RA model would not be appropriate. In this model, the A parameter reflects a late automatic process that influences behavior only when control has failed, and the model is not mathematically equipped to estimate early automatic processes. Similarly, if a researcher is interested in a controlled process that overcomes the influence of such automatic processes, then neither the RA model nor the ID model would be appropriate. Neither one of those models is mathematically equipped to estimate such overcoming bias processes, and neither one of those models is able to separate the extent of activation from the extent of overcoming that activation (e.g., identical Stroop performance from an adult and a child). In the same way, if a researcher is interested in measuring self-regulatory processes that are not conditional on association activation and detection (see Footnote 3), then the Quad model would not be appropriate. Thus, in selecting a model, a paramount concern should be which processes are theoretically relevant to the research. Furthermore, having applied a particular model, researchers must clearly understand exactly what types of automatic and controlled processes the models estimate, and researchers should interpret their results accordingly.

One important point to emphasize in this regard is that the automatic and controlled processes estimated by the PD models and the Quad model are not identical. The controlled processes estimated with the two PD models (early control, late control) are different from one another and are both different from the detection (D) and overcoming bias (OB) parameters estimated with the Quad model. In the RA model, operation of the early control process is completely unaffected by automatic processes, and the early control process drives behavior entirely on its own. In the ID model, the late control process cannot be engaged if the automatic process operates. In contrast, the D parameter of the Quad model always influences judgments, regardless of the status of the other parameters, and does not preclude the operation of automatic processes. Similarly, the OB parameter is not precluded by automatic processes, nor does it, in turn, preclude automatic processes. Indeed, the operation of OB is dependent on both association activation (AC) and D.

The automatic processes estimated by the two PD models (late bias, attention capture) also are different from one another and are both different from the AC and guessing (G) parameters estimated by the Quad model. In the RA model, the late automatic process operates only in the wake of failed control. In the ID model, operation of the automatic process precludes controlled processing and directs behavior entirely on its own. In contrast, the AC parameter of the Quad model always influences judgments, regardless of the status of the other parameters, and does not preclude the operation of controlled processes. The G parameter operates only in the absence of all other processes and is thus dependent on the failure of both controlled (D) and automatic (AC) processes.

The important point here is that the Quad model does not simply represent an extension of PD models that includes the same processes as those models with a few additional ones. Rather, the processes represented by each of the models are fundamentally different. In the same way one may choose to measure attention capture, attentional disengagement, perceptual encoding, conceptual encoding, or any number of other processes in standard

behavioral research, when choosing a model, one must decide which processes are most important to measure. The parameters estimated by each model are, in fact, separate dependent variables representing different, specific cognitive processes. In all cases, the selection of an appropriate dependent variable depends on the research question.

*Predictive validity.* A second basis for choosing among models is the extent to which the models' parameters have been shown to predict behavior. As described above, the parameters of the Quad model have been shown to be related to a variety of contextual variables and interventions that influence implicit bias. The parameters also predict important group, individual, and developmental differences in implicit bias. Finally, the parameters have been shown to predict important extra-task behavior related to interpersonal interactions.

*Model fit.* A final important basis for choosing among models is model fit. All else being equal, the model that provides the best account of the data is preferred. However, in comparing the fits of different models, it is important to account for the complexity of the models because more complex models tend to fit given data better than do simpler models. For example, because the Quad model estimates four parameters, rather than the two parameters estimated by PD models, the Quad model will tend to provide superior model fit. As such, an important goal is to find the best compromise between fit and parsimony. To do so, one should use selection criteria that penalize models for complexity. Akaike information criterion (AIC) and Bayes information criterion (BIC) are two metrics of model fit that correct for model complexity (for a review, see Myung, 2000).

We undertook an extensive comparison of the fits of the Quad model and the RA model. We chose this comparison because the RA model is the PD model that Payne (e.g., Payne, 2001) and his colleagues have advocated as the appropriate one to model implicit measures of attitudes and beliefs, particularly priming tasks, such as the weapon identification task. For a total of 57 data sets (50 IAT data sets; 7 priming data sets), we generated estimates of AIC and BIC for both the Quad model and the RA model. Overall, the Quad model provided significantly better AIC,  $t(56) = 5.95$ ,  $p < .01$ , and BIC,  $t(56) = 2.85$ ,  $p < .01$ , than did the RA model. However, each model provided better fit for one type of measure. For IAT studies, the Quad model provided significantly better AIC,  $t(49) = 6.37$ ,  $p < .01$ , and BIC,  $t(49) = 3.30$ ,  $p < .01$ , than did the RA model.<sup>7</sup> In contrast, for priming studies, the RA model provided marginally better AIC,  $t(6) = 2.33$ ,  $p < .06$ , and significantly better BIC,  $t(6) = 4.95$ ,  $p < .01$ , than did the Quad model.

From these data, one might be tempted to conclude that the Quad model should be used to model IAT data, whereas the RA model should be used to model priming data. However, the effect sizes of the differences in AIC and BIC were very small in all comparisons (Cohen's  $d$ s  $< .017$ ). As such, in our view, either model may be applied to either task, provided that the model provides adequate fit for the given data set. Given that both models can accurately account for the data, and given the small differences in AIC and BIC, we once again suggest that the primary consid-

<sup>7</sup> The Quad model also produced substantially better fit than did the RA model for the one GNAT data set that was collected.



eration should be theoretical. After all, the choice between conceptually distinct measures should be determined a priori by the theoretical research question, rather than a posteriori by the outcome these measures produced. If one wishes to measure an early control process that precludes automatic influences, then one should use the RA model, regardless of whether the Quad model provides slightly superior model fit. If one wishes to examine the self-regulation of automatic associations (the OB parameter), then one should use the Quad model, regardless of whether the RA model provides slightly superior fit. Given that both models and their parameters have been extensively validated, they may each be used to learn different things from the same data set.

### *Dual System Models*

Recently, a number of generalized dual-process models have been proposed that aim to provide comprehensive integrations across domain-specific dual-process models (e.g., Gawronski & Bodenhausen, 2006; Lieberman, Gaunt, Gilbert, & Trope, 2002; Smith & DeCoster, 2000; Strack & Deutsch, 2004). These models are not concerned with the specifics of how different types of automatic and controlled processes might be measured but rather are concerned with providing a broad framework for conceptualizing the distinction across content domains and measures. The integrative nature of these models is their greatest strength. Another appealing feature is that they each tie different types of processes to particular neural structures. In so doing, the models establish important links to research on functional neuroanatomy that can inform our understanding of the interaction of multiple, qualitatively distinct processes.

The integrative approach of these models is compatible with that of the Quad model. However, there also are some important differences among the models (for a review, see Sherman, 2006a, 2006b). The most practical (and perhaps important) difference between the Quad model and the dual-system models is that the Quad model provides a means of quantifying the independent and simultaneous contributions of distinct processes within a single task. There also are conceptual differences among the models. For example, Smith and DeCoster's (2000) model is not concerned with the type of self-regulatory control represented by the overcoming bias parameter in the Quad model. Lieberman et al.'s (2002) model proposes that controlled processes are engaged primarily when automatic processes do not determine behavior (though the controlled process may then feed back into and influence the automatic process). Thus, the model is generally not concerned with cases in which automatic influences operate only in the wake of failed control (e.g., Jacoby, Kelley, & McElree, 1999).

The data we have collected in support of the Quad model place constraints on what can be said about separate processing systems. Strong statements that different measures are affected exclusively by one or the other system (e.g., Rydell & McConnell, 2006) have trouble accounting for our findings that multiple automatic and controlled processes contribute to implicit task performance. Just as implicit and explicit measures should not be taken as direct reflections of automatic and controlled processes, they also should not be taken as proxies for separate systems of reasoning or evaluation.

Our data also raise important questions about the operation and definition of processing systems. All of the dual system models propose two systems, each of which includes a wide variety of different processes. Thus, the controlled, reflective, or rule-based system includes many different controlled processes, and the automatic, impulsive, or associative system includes many different automatic processes. However, as described in our discussion of different automatic and controlled processes, there are many behavioral and neurological dissociations among these processes. Thus, the categories automatic and controlled are highly heterogeneous. Consistent with this depiction, our tests of the Quad model have demonstrated that contextual factors and interventions may influence two different automatic or controlled processes independently or even in opposite directions. For example, our reanalysis of Lambert et al.'s (2003) data showed that a public context simultaneously decreased the ability to detect correct responses and increased the ability to overcome automatic biases. Likewise, individual differences may be differentially related to different automatic and controlled processes. For example, aging was associated with increases in the ability to detect correct responses but decreases in the ability to overcome automatic biases. Such findings are problematic for system models that propose the operation of unitary systems to which multiple automatic and controlled processes belong. The necessary implication is that one system is simultaneously producing two opposing effects.

This raises the issue that at some point the basic level (e.g., Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) of a system needs to be defined. How many distinct processes can belong to a single system? What are the features that make a collection of processes a system? One must settle on some balance of breadth and specificity when enumerating different categories or types of processes, and this balance may be context-dependent. In the Quad model, we make a rough distinction among four different processes and argue that it is useful to assess all four within a given behavior. However, in some cases, the simple distinction found in dual-process models between automatic and controlled processes may well suffice. In other cases, researchers may wish to further divide automatic and controlled processes into six or eight key components. As we have noted earlier, the four processes elaborated in the Quad model are certainly not an exhaustive account of all possible, relevant processes.

### SUMMARY

In this article, we described the Quad model, an account of how people regulate automatic associations and behavioral impulses. The model proposes the joint operation of four processes to achieve such self-regulation. In specifying four qualitatively distinct processes, the Quad model integrates and expands on the many dual-process models of psychological processing. It is important to note that the model is implemented as a multinomial model that provides independent quantitative estimates of each of the processes. We described research that validates the model and its parameters. We also described applications of the model to a number of important theoretical issues relating to the operation of automatic and controlled processes. These issues include the basic

nature and meaning of implicit measures of association, the bases of contextual variability of implicit bias, the effects of interventions to change these biases, and the bases of individual, group, and developmental variations in these biases. Our research shows that for each of these topics the influence of controlled processes has been substantially underestimated. We also presented evidence that the model's parameters can be used to account for broad, domain-relevant behavior outside the immediate context from which the parameters were derived. Thus, the Quad model is a general model of self-regulation that may be applied to any situation in which automatic associations or impulses conflict with more controlled processes. Such conflicts are relevant to many important domains of psychology, including those related to addictions, phobias, habits, dieting, aggression, emotion, decision making, and many other self-control dilemmas. In each of these domains, the Quad model can help answer important questions about what the determinants of behavior are, what the bases of individual differences in behavioral outcomes are, and how behavior changes as a function of interventions, context, and development.

## References

- Allen, T. J., Sherman, J. W., & Gonsalkorale, K. (2008). [The underlying bases of contextual variation in implicit bias]. Unpublished data.
- Amodio, D. M., Devine, P. D., & Harmon-Jones, E. (2008). Individual differences in the regulation of race bias: The role of conflict detection and neural signs for control. *Journal of Personality and Social Psychology*, 94, 60–74.
- Amodio, D. M., Harmon-Jones, E., & Devine, P. G. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *Journal of Personality and Social Psychology*, 84, 738–753.
- Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between explicit and implicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology*, 83, 380–393.
- Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology*, 81, 789–799.
- Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology*, 68, 181–198.
- Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology*, 87, 5–22.
- Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition: Vol. 1. Basic Processes* (2nd ed., pp. 1–40). Hillsdale, NJ: Erlbaum.
- Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). New York: Guilford Press.
- Baron, R. S. (1986). Distraction-conflict theory: Progress and problems. *Advances in Experimental Social Psychology*, 19, 1–40.
- Bartholow, B. D., Dickter, C. L., & Sestir, M. A. (2006). Stereotype activation and control of race bias: Cognitive control of inhibition and its impairment by alcohol. *Journal of Personality and Social Psychology*, 90, 272–287.
- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin and Review*, 6, 57–86.
- Baumeister, R. F., & Vohs, K. D. (2004). *Handbook of self-regulation: Research, theory, and applications*. New York: Guilford Press.
- Beck, A. T., & Clark, D. A. (1997). An information processing model of anxiety: Automatic and strategic processes. *Behaviour Research and Therapy*, 35, 49–58.
- Beer, J. S., Shimamura, A. P., & Knight, R. T. (2004). Frontal lobe contributions to executive control of cognitive and social behavior. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (3rd ed., pp. 1091–1104). Cambridge, MA: MIT Press.
- Beer, J. S., Stallen, M., Lombardo, M. V., Gonsalkorale, K., Cunningham, W. A., & Sherman, J. W. (2007). *Neural underpinnings of implicit prejudice: The Quadruple process model approach*. Unpublished manuscript.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6, 242–261.
- Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, 81, 828–841.
- Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999, November 11). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, 402, 179–181.
- Bowles, B., Crupi, C., Mirsattari, S. M., Pigott, S. E., Parrent, A. G., Pruessner, J. C., et al. (2007). Impaired familiarity with preserved recollection after anterior temporal-lobe resection that spares the hippocampus. *Proceedings of the National Academy of Sciences, USA*, 104, 16382–16387.
- Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Erlbaum.
- Brewer, M. B., & Feinstein, A. S. H. (1999). Dual processes in the cognitive representations of persons and social categories. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 255–270). New York: Guilford.
- Britton, J. C., Taylor, S. F., Sudheimer, K. D., & Liberzon, I. (2006). Facial expressions and complex IAPS pictures: Common and differential networks. *NeuroImage*, 31, 906–919.
- Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control-theory approach to human behavior*. New York: Springer-Verlag.
- Chaiken, S. (1980). Heuristic versus systematic processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39, 752–766.
- Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Chee, M. L., Sriram, N., Soon, C. S., & Lee, K. M. (2000). Dorsolateral prefrontal cortex and the implicit association of concepts and attributes. *Neuroreport for Rapid Communication of Neuroscience Research*, 11, 135–140.
- Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73–96). New York: Guilford Press.
- 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.
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83, 1314–1329.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and

- disliked individuals. *Journal of Personality and Social Psychology*, 81, 800–814.
- De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology*, 50, 77–85.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18.
- Devine, P. G., Plant, E. A., 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.
- Easdon, C. M., & Vogel-Sprott, M. (2000). Alcohol and behavioral control: Impaired response inhibition and flexibility in social drinkers. *Experimental and Clinical Psychopharmacology*, 8, 387–394.
- Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, 48, 269–297.
- Ellwart, T., Becker, E. S., & Rinck, M. (2005). Activation and measurement of threat associations in fear of spiders: An application of the extrinsic affective Simon task. *Journal of Behavior Therapy and Experimental Psychiatry*, 36, 281–299.
- Entman, R. M., & Rojecki, A. (2001). *The Black image in the White mind: Media and race in America*. Chicago: University of Chicago Press.
- Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist*, 49, 709–724.
- Erb, H., Kruglanski, A. W., Chun, W. Y., Pierro, A., Mannetti, L., & Spiegel, S. (2003). Searching for commonalities in human judgment: The parametric unimodel and its dual model alternatives. *European Review of Social Psychology*, 14, 1–49.
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology*, 23, 75–109.
- 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.
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229–238.
- Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 97–116). New York: Guilford.
- Ferguson, M. J., & Bargh, J. A. (2004). Liking is for doing: The effects of goal pursuit on automatic evaluation. *Journal of Personality and Social Psychology*, 87, 557–572.
- Ferguson, M. J., & Bargh, J. A. (2007). Beyond the attitude object: Implicit attitudes spring from object-centered contexts. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies* (pp. 216–246). New York: Guilford Press.
- 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.
- Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 231–254). New York: Guilford.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, 23, 1–74.
- Fujita, K., Trope, Y., Liberman, N., & Levin-Sagi, M. (2006). Construal levels and self-control. *Journal of Personality and Social Psychology*, 90, 351–367.
- Gawronski, B., & Bodenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation and retrieval experiences. *Journal of Personality and Social Psychology*, 89, 672–685.
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, 132, 692–731.
- Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When “just say no” is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology*, 44, 370–377.
- Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, 2, 181–193.
- Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford Press.
- Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, 54, 733–740.
- Glaser, J., & Knowles, E. D. (2008). Implicit motivation to control prejudice. *Journal of Experimental Social Psychology*, 44, 164–172.
- Gonsalkorale, K., & Sherman, J. W. (2007). *Altering newly formed implicit attitudes: An examination of the underlying mechanisms*. Unpublished raw data.
- Gonsalkorale, K., Sherman, J. W., Allen, T. J., Amodio, D. M., & Bartholow, B. D. (2007). *On the underlying bases of implicit attitude variability and malleability*. Unpublished manuscript.
- Gonsalkorale, K., von Hippel, W., & Sherman, J. W. (2007). *Bias and regulation of bias in intergroup interactions*. Unpublished manuscript.
- Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology*, 40, 357–365.
- Gray, N. S., McCulloch, M. J., Smith, J., Morris, M., & Snowden, R. J. (2003, May 29). Violence viewed by psychopathic murderers: Adapting a revealing test may expose those psychopaths who are most likely to kill. *Nature*, 423, 497–498.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480.
- Greenwald, A. G., Oakes, M. A., & Hoffman, H. G. (2003). Targets of discrimination: Effects of race on responses to weapons holders. *Journal of Experimental Social Psychology*, 39, 399–405.
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.
- Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, 53, 575–604.
- Huijding, J., & de Jong, P. J. (2005). A pictorial version of the extrinsic affective Simon task: Sensitivity to generally affective and phobia-relevant stimuli in high and low spider fearful individuals. *Experimental Psychology*, 52, 289–295.
- Huijding, J., & de Jong, P. J. (2006). Specific predictive power of automatic spider-related affective associations for controllable and uncontrollable fear responses toward spiders. *Behaviour Research and Therapy*, 44, 161–176.
- Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. New York: Appleton-Century.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513–541.
- Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 3–22.



- Jacoby, L. L., Begg, I. M., & Toth, J. P. (1997). In defense of functional independence: Violations of assumptions underlying the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 484–495.
- Jacoby, L. L., Bishara, A. J., Hessels, S., & Toth, J. P. (2005). Aging, subjective experience, and cognitive control: Dramatic false remembering by older adults. *Journal of Experimental Psychology: General*, 134, 131–148.
- Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: Early selection versus late correction. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 383–402). New York: Guilford Press.
- Jacoby, L. L., McElree, B., & Trainham, T. N. (1999). Automatic influences as accessibility bias in memory and Stroop-like tasks: Toward a formal model. In A. Koriath & D. Gopher (Eds.), *Attention and performance XVII: Cognitive regulation of performance: interaction of theory and application* (pp. 461–486). Cambridge, MA: MIT Press.
- Jacoby, L. L., Shimizu, Y., Daniels, K. D., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin and Review*, 12, 852–857.
- Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, 122, 139–154.
- Jonides, J., Sylvester, C. C., Lacey, S. C., Wager, T. D., Nichols, T. E., & Awh, E. (2003). Modules of working memory. In R. H. Kluwe, G. Lüer, & F. Rösler (Eds.), *Principles of learning and memory* (pp. 113–134). Cambridge, MA: Birkhäuser.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58, 697–720.
- Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, 78, 871–888.
- Kim, D. Y. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly*, 66, 83–96.
- Kim, D., Sarason, B. R., & Sarason, I. G. (2006). Implicit social cognition and culture: Explicit and implicit psychological acculturation, and distress of Korean-American young adults. *Journal of Social and Clinical Psychology*, 25, 1–32.
- Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7–49). Mahwah, NJ: Erlbaum.
- Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the “Who said what?” paradigm. *Journal of Personality and Social Psychology*, 75, 1155–1178.
- Krull, D. S., & Dill, J. C. (1996). On thinking first and responding fast: Flexibility in social inference processes. *Personality and Social Psychology Bulletin*, 22, 949–959.
- Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, 103, 284–308.
- Lambert, A. J., Cronen, S., Chasteen, A. L., & Lickel, B. (1996). Private versus public expressions of racial prejudice. *Journal of Experimental Social Psychology*, 32, 437–459.
- Lambert, A. J., Payne, B. K., Jacoby, L. L., Shaffer, L. M., Chasteen, A. L., & Khan, S. R. (2003). Stereotypes as dominant responses: On the “social facilitation” of prejudice in anticipated public contexts. *Journal of Personality and Social Psychology*, 84, 277–295.
- Lazarus, R. S. (1982). Thoughts on the relation between emotion and cognition. *American Psychologist*, 37, 1019–1024.
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125, 255–275.
- Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology*, 34, 199–249.
- Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 219–234.
- Loewenstein, G. F. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, 65, 272–292.
- Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, 81, 842–855.
- 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.
- Maison, D., Greenwald, A. G., & Bruin, R. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology*, 14, 405–415.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87, 252–271.
- Mathews, A., & MacLeod, C. (1994). Cognitive approaches to emotion and emotional disorders. *Annual Review of Psychology*, 45, 25–50.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99, 440–466.
- Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification. *Psychological Review*, 106, 3–19.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.
- Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, 132, 455–469.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100.
- 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.
- Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, 132, 297–326.
- Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology*, 77, 167–184.
- Moskowitz, G. B., Salomon, A. R., & Taylor, C. M. (2000). Preconsciously controlling stereotyping: Implicitly activated egalitarian goals prevent the activation of stereotypes. *Social Cognition*, 18, 151–177.
- Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, 126, 247–259.
- Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: A meta-analysis. *Cognitive, Affective, and Behavioral Neuroscience*, 3, 207–233.
- Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, 44, 190–204.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, 19, 625–666.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting

- implicit group attitudes and beliefs from a demonstration website. *Group Dynamics*, 6, 101–115.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31, 166–180.
- Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81, 181–192.
- Payne, B. K., Burkley, M., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, 94, 16–31.
- Payne, B. K., Lambert, A. J., & Jacoby, L. L. (2002). Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. *Journal of Experimental Social Psychology*, 38, 384–396.
- Perdue, C. W., Dovidio, J. F., Gurtman, M. B., & Tyler, R. B. (1990). Us and them: Social categorization and the process of intergroup bias. *Journal of Personality and Social Psychology*, 59, 475–486.
- Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. Dubuque, IA: Brown.
- Petty, R. E., & Wegener, D. T. (1999). The elaboration likelihood model: Current status and controversies. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 41–72). New York: Guilford Press.
- Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, 16, 331–348.
- Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, 12, 729–738.
- Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.
- Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology*, 39, 475–573.
- Richeson, J. A., & Ambady, N. (2003). Effects of situational power on automatic racial prejudice. *Journal of Experimental Social Psychology*, 39, 177–183.
- Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, 95, 318–339.
- Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist*, 9, 1043–1056.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382–439.
- Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. K. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition*, 17, 437–465.
- Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Development*, 76, 451–466.
- Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, 91, 995–1008.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1–66.
- Sechrist, G. B., & Stangor, C. (2001). Perceived consensus influences intergroup behavior and stereotype accessibility. *Journal of Personality and Social Psychology*, 80, 645–654.
- Seibt, B., Häfner, M., & Deutsch, R. (2007). Prepared to eat: How immediate affective and motivational responses to food cues are influenced by food deprivation. *European Journal of Social Psychology*, 37, 359–379.
- Sherman, J. W. (2006a). Clearing up some misconceptions about the Quad model. *Psychological Inquiry*, 17, 269–276.
- Sherman, J. W. (2006b). 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., Groom, C. J., Ehrenberg, K., & Klauer, K. C. (2003). Bearing false witness under pressure: Implicit and explicit components of stereotype-driven memory distortions. *Social Cognition*, 21, 213–246.
- Sherman, S. J., Rose, J. S., Koch, K., Presson, C. C., & Chassin, L. (2003). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology*, 22, 13–39.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127–190.
- Simonton, D. K. (1998). Achieved eminence in minority and majority cultures: Convergence versus divergence in the assessments of 294 African Americans. *Journal of Personality and Social Psychology*, 74, 804–817.
- Sinclair, S., Lowery, B. S., Hardin, C. D., & Colangelo, A. (2005). Social tuning of automatic racial attitudes: The role of affiliative motivation. *Journal of Personality and Social Psychology*, 89, 583–592.
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22.
- Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4, 108–131.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8, 220–247.
- Stroop, J. R. (1935). Studies on the interference in serial verbal reactions. *Journal of Experimental Psychology*, 59, 239–245.
- Taylor, S. F., Kornblum, S., Lauber, E. J., Minoshima, S., & Koeppe, R. A. (1998). Isolation of specific interference processing in the Stroop task: PET Activation studies. *Journal of Neuroimaging*, 6, 81–92.
- Teachman, B. A. (2007). Evaluating implicit spider fear associations using the go/no-go association task. *Journal of Behavior Therapy and Experimental Psychiatry*, 38, 156–167.
- Teachman, B. A., Gregg, A. P., & Woody, S. R. (2001). Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology*, 110, 226–235.
- Teachman, B. A., & Woody, S. R. (2003). Automatic processing in spider phobia: Implicit fear associations over the course of treatment. *Journal of Abnormal Psychology*, 112, 100–109.
- Toth, J. P., Reingold, E. M., & Jacoby, L. L. (1994). Toward a redefinition of implicit memory: Process dissociations following elaborative processing and self-generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 290–303.
- Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, 93, 239–257.
- Trope, Y., & Fishbach, A. (2000). Counteractive self-control in overcoming temptation. *Journal of Personality and Social Psychology*, 79, 493–506.
- Trope, Y., & Gaunt, R. (2000). Processing alternative explanations of behavior: Correction or integration? *Journal of Personality and Social Psychology*, 79, 344–354.
- Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1130.
- von Hippel, W., Silver, L. A., & Lynch, M. E. (2000). Stereotyping against your will: The role of inhibitory ability in stereotyping and prejudice



- among the elderly. *Personality and Social Psychology Bulletin*, 26, 523–532.
- Wagner, A. D., Stebbins, G. T., Masciari, F., Fleischman, D. A., & Gabrieli, J. D. E. (1998). Neuropsychological dissociation between recognition familiarity and perceptual priming in visual long-term memory. *Cortex*, 34, 493–511.
- Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, 101, 34–52.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107, 101–126.
- Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, 72, 262–274.
- 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.
- Zajonc, R. B. (1965, July 16). Social facilitation. *Science*, 149, 269–274.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151–175.

Received January 30, 2004

Revision received December 19, 2007

Accepted December 20, 2007 ■