Title

Controlled influences on implicit measures: Confronting the myth of process-purity and taming the cognitive monster

Author
Sherman, J

Publication Date
2023-12-10

Peer reviewed

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45*, 9, 1043–1056.
Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test (IAT): Dissociating salience from associations. *Journal of Experimental Psychology: General, 133*, 139–165.
Schnabel, K., Banse, R., & Asendorpf, J. B. (2006). Employing automatic approach and avoidance tendencies for the assessment of implicit personality self-concept: The Implicit Association Procedure (IAP). *Experimental Psychology, 53*, 69–76.
Steffens, M. C. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology, 51*, 165–179.
Steffens, M. C., & Plewe, I. (2001). Items' cross-category associations as a confounding factor in the Implicit Association Test. *Zeitschrift für Experimentelle Psychologie, 48*, 123–134.
Teige, S., Schnabel, K., Banse, R., & Asendorpf, J. B. (2004). Assessment of multiple implicit self-concept dimensions using the Extrinsic Affective Simon Task (EAST). *European Journal of Personality, 18*, 495–520.
Vargas, P. T. (2004). On the relationship between implicit attitudes and behavior: Some lessons from the past, and directions for the future. In G. Haddock & G. R. Maio (Eds.), *Contemporary perspectives on the psychology of attitudes*. New York: Psychology Press.
Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101–126.
Zajonc, R. B. (1980). Feeling and thinking. Preferences need no inferences. *American Psychologist, 35*, 151–175.

13

Controlled Influences on Implicit Measures*

Confronting the Myth of Process-Purity and Taming the Cognitive Monster

Jeffrey W. Sherman

Introduction

Though implicit measures often are portrayed as *process-pure* measures of automatic attitudes, instead, they reflect the joint contributions of automatic and controlled processes. As such, automatic and controlled components of attitudes are better measured, not with two separate measures, but with process dissociation (PD) techniques that extract independent estimates of automatic and controlled influences from performance on a single task. I will describe such approaches for analyzing responses on implicit tasks, concentrating on our own Quadruple Process (Quad) model (Sherman, Gawronski, Hugenburg, & Groom, 2005; Sherman, Gawronski, Conrey, Hugenburg, & Groom, 2008). The application of the Quad model provides important insights into central questions surrounding the conceptualization, measurement, and interpretation of implicit attitudes. Among these, Quad model analyses shed new light on Bargh's (1999) provocative claim that the "cognitive monsters" of implicit attitudes and stereotypes cannot be controlled. I will reexamine the status of this claim, concluding that, in general, the controllability of such biases has been underestimated. However, some forms of control may be more attainable than others. In particular, I

* This research was supported by NIMH grant R01 MH59774.

will argue that a failure to consider the multifaceted nature of implicit task performance has led to both an overestimation of the ease with which people can control the initial automatic activation of relevant associations and an underestimation of the extent to which the expression of those associations subsequently can be controlled.*

The Myth of Process-Purity and the Trouble with Tasks

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 of attitudes has offered promising new ways to avoid these obstacles by measuring attitudes without directly requesting that respondents report those attitudes. 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 attitudes, whose expression largely cannot be altered or inhibited (e.g., Bargh, 1999; Devine, 1989; Fazio, Jackson, Dunton, & Williams, 1995; Greenwald, McGhee, & Schwartz, 1998; Kim, 2003). Taken

* In this chapter, I will use the term *implicit measure* to refer to measures that assess attitudes and knowledge indirectly (i.e., without explicitly asking people to report their attitudes and knowledge). The term *indirect measure* may be technically more accurate for my intended meaning, but I will nevertheless use the common terminology of *implicit*.

I will use the term *implicit attitude* to refer simply to an attitude that is measured with an implicit measure. Here, the term implicit attitude implies nothing about the status of that attitude in terms of subjective awareness, intention, controllability, and so on. In my view, those are empirical questions that may be asked about any measured attitude, rather than definitional criteria. Critically, though implicit attitudes are defined as the behavioral outcomes of implicit measures (e.g., reaction time effects), they are not assumed to be isomorphic with the underlying evaluative associations (or *evaluative generators*) that instigate responses on the measures. Rather, behavioral biases on implicit measures (i.e., implicit attitudes) may or may not correspond closely with underlying associations, depending on the intervention of other processes that translate the associations into behavioral responses on the implicit measures. Thus, I call the behavioral bias an implicit attitude in the common vernacular, but distinguish this attitude from evaluative associations.

† The IAT may be a notable exception to this state of affairs (e.g., Fazio & Olson, 2003; Monteith, Voils, & Ashburn-Nardo, 2001).

in conjunction with explicit measures (e.g., questionnaires), implicit measures are used to compare and contrast automatic and controlled aspects of attitudes.

The Specter of Task Confounds

Though this *task dissociation* approach has certainly proven to be productive (see the contents of this book), it has significant drawbacks. First, assessing automatic and controlled components of attitudes with separate measures introduces a confound between process type (automatic vs. controlled) and measurement task (e.g., IAT vs. questionnaire). Undoubtedly, implicit measures are less subject to the whims of awareness and intention than are explicit measures. However, there may be other important differences between any pair of implicit and explicit tasks beyond the extent to which they tap automatic versus controlled processing.

As an example, consider the case of research on implicit and explicit memory. As in the attitudes domain, for years, different measures were used to assess what were thought to be independent implicit and explicit types or systems of memory. However, Roediger and his colleagues determined that, whereas implicit measures of memory had tapped perceptual encoding processes, explicit measures had tapped conceptual encoding processes (e.g., Roediger, 1990). Instantaneously, a whole generation of research depicting differences between implicit and explicit types of memory was open to reinterpretation as reflecting, instead, differences in measures that tapped perceptual and conceptual encoding processes. As of yet, no one has provided a similar reinterpretation of dissociations between implicit and explicit measures of attitudes. However, Roediger's example should serve as a cautionary tale for social psychologists' approach to studying implicit attitudes. Even if there is no overarching confound across all implicit and explicit measures of attitudes, within any pair of measures, there are bound to be significant differences beyond the extent to which they tap automatic and controlled processes. Thus, assessing automatic and controlled aspects of attitudes via task dissociations is problematic.

On Process-Purity

The more general point is that no task is process-pure. Any task that requires an observable response (e.g., a button press) cannot be entirely automatic, and no task is immune from the influence of automatic processes (e.g., Jacoby, Toth, & Yonelinas, 1993). Rather, all tasks involve an ongoing interplay among simultaneously occurring automatic and controlled processes. As such, a behavioral response, in and of itself, is incapable of specifying the nature of the underlying processes that produced the response.

Consider the Stroop Task (Stroop, 1935). A fully literate adult and a young child who knows colors but does not know how to read may make an equally small number of errors on the task. However, very different processes are at work for the adult and the child. On incompatible trials (e.g., the word *blue* written in red ink), the adult must overcome a habit to read the word in order to name the color of the ink correctly. In contrast, the child has no habit to overcome; she or he simply responds to the color of the ink.

The same principle applies to implicit measures of attitudes, many of which have a Stroop-like structure of compatible (e.g., pairing Black faces with negative words and White faces with positive words) and incompatible (e.g., pairing Black faces with positive words and White faces with negative words) trials. The performance of two people who appear to have equally strong implicit attitudes on such measures may reflect very different underlying processes. Whereas one person may have strong automatic evaluative associations that are successfully overcome in responding, the other may have weaker associations that are not overcome so well. The measure itself cannot distinguish between the two cases. The distinction is well worth making because the causes, consequences, and cures of having strong automatic associations versus weak self-regulatory abilities are very different.

An important methodological implication of this analysis is that, when taken as pure reflections of automatic associations, implicit measures underestimate the extent of cognitive control. The equally important corollary is that a strong ability to overcome automatic associations on implicit measures may mask the true extent of automatic bias (e.g., Conrey et al., 2005; Sherman et al., 2008).

The Poverty of Task Dissociation

A related drawback to the task dissociation approach is that it cannot reveal the simultaneous contributions of automatic and controlled processes to attitudinal responses. If we assume that responses on any attitude measure reflect the joint contributions of automatic and

controlled processes, then it would be advantageous to have a means to track those contributions independently. However, because implicit and explicit measures are taken as self-contained, process-pure estimates of automatic and controlled processes, there is no way to assess the ongoing interplay of these processes in producing a discrete attitudinal response on a particular task. This necessarily produces an overly simplified depiction of the processes that underlie the production of attitudinal responses.

Implicit Attitudes are Constructed, Not Revealed

The preceding discussions all converge on the important point that responses on implicit measures of attitudes are just that: responses on measures. As such, there are all sorts of factors and processes that may intervene in the translation of evaluative associations into responses on implicit measures. Though the constructive nature of responses on explicit measures of attitudes has been well appreciated (e.g., Wilson & Hodges, 1992), the same has not been true for implicit measures. Because responses on implicit measures are typically viewed as inevitable and uncontrollable (e.g., Bargh, 1999; Devine, 1989), they have been portrayed as reflecting a real, true, and singular underlying representation to a much greater extent than have responses on explicit tasks (e.g., Dovidio & Fazio, 1992; Fazio et al., 1995). It is as if implicit measures crawl inside our heads and locate a treasure chest that contains the One Real Attitude, revealing the buried truth that people may be unable or unwilling to report.

However, though implicit measures are certainly less susceptible to intention and less reliant on awareness than are explicit measures, recent evidence makes clear that implicit attitudes are not the singular, stable entities they once were thought to be. For example, there is now considerable evidence that responses on implicit measures may be influenced by a variety of personal and contextual factors (e.g., Blair, 2002). Moreover, implicit measures of attitudes show poorer test-retest reliability than do explicit measures and show smaller correlations across measures of the same attitude object than do explicit measures (e.g., Cunningham, Preacher, & Banaji, 2001; Kawakami & Dovidio, 2001). These findings are hard to reconcile with the view that implicit measures directly tap singular, true attitudes. Instead, these results indicate that implicit measures are no different than all other psychological measures; there is a translational gap between the construct and the way it is measured.

There are many possible explanations for the apparently large size of this gap in implicit measures of attitudes. To be sure, random measurement error is one important contributor to the lack of reliability and convergent validity among implicit measures (e.g., Cunningham et al., 2001). The lack of convergent validity also likely reflects structural differences in the extent to which different measures tap attitudes toward particular category exemplars versus categories as a whole (e.g., Fazio & Olson, 2003; Olson & Fazio, 2003). In this case, the lack of validity does not represent the failure of two measures of the same attitude to correlate, but rather the fact that the two measures simply tap different attitudes.

However, implicit measures differ not only in the content that they tap, but also in the processes that they recruit for task completion. That is, different tasks (e.g., IAT, evaluative priming, Weapons Identification Task, or WIT; Shooter Task), and even different stimuli within a single task, recruit different processes to differing degrees. For example, implicit tasks may differ widely in the extent to which they require accurate perception of a stimulus or are susceptible to regulatory efforts at overcoming automatically activated associations (e.g., Conrey et al., 2005; Sherman et al., 2008). As such, low levels of reliability and construct validity in implicit measures reflect more than random noise in the measurement of a true, singular attitude. They also reflect more than the fact that different implicit measures may tap different singular, true attitudes. Even if random noise and activated content can be controlled, the fact remains that different measures of the exact same content will vary in the procedural demands they make on the respondent. Once again, implicit attitudes are constructed in response to the demands of the task; they are not revealed truths.

This is not meant to suggest that there is no such thing as a stable evaluation that is represented in memory. In any given case, the extent to which responses on an implicit measure are influenced by strong and stable evaluative or descriptive associations will vary (e.g., Conrey et al., 2005; Sherman et al., 2008). The point is that these underlying associations must be translated into what we call implicit attitudes via performance on some task. There are significant drawbacks to treating these task outcomes as direct reflections of automatically activated and uncontrollable associations, rather than as constructed responses that reflect the influence of multiple component processes. These drawbacks are particularly consequential when implicit and explicit measures

are compared for the purpose of drawing conclusions about the automatic versus controlled nature of attitudes. As summarized above, this approach increases the risk of misinterpreting task confounds, increases the risk of misrepresenting the extent of automatic and controlled influences on an attitude, and thwarts the identification of joint, ongoing automatic and controlled processes.

Process Dissociation

These same concerns about task dissociation paradigms in the implicit memory literature led Jacoby (1991; Lindsay & Jacoby, 1994; see Jacoby, Kelley, & McElree, 1999, for a review) to develop process dissociation (PD) techniques for separating the automatic and controlled components of behavior from performance on a single task. The PD approach assumes that no measure is process-pure, and that automatic and controlled processes exert independent and simultaneous influences on any task. Because estimates of the two components are derived from behavior on a single task, PD techniques avoid confounding task and process.

Initially, Jacoby and his colleagues developed two different models of process dissociation (Jacoby, 1991; Lindsay & Jacoby, 1994). Both models rely on the method of opposition, which contrasts performance on compatible trials, on which a controlled process and an automatic process should lead to the same response, with performance on incompatible trials, on which the two processes should lead to different responses. Estimates of automatic and controlled processing are derived by subtracting performance on the incompatible trials from performance on the compatible trials. The primary difference between the two models is whether automatic or controlled processes are assumed to be primary.

The Control Default Model

One model is designed to account for tasks in which automatic processes are thought to influence behavior only when control fails (Jacoby, 1991). For example, in recognition memory, Jacoby proposed that controlled, effortful recollective processes will determine judgments whenever possible. Only when controlled recollection fails to provide a response will automatically generated perceptions of an item's familiarity drive recognition judgments. Estimates of the contributions of recollection (control) and familiarity (automaticity) are derived by comparing trials in which the two processes produce the same recognition judgment with trials in which the two processes lead to different recognition judgments (e.g., an item that is familiar but was not presented as part of the to-be-remembered list). This model of process dissociation is frequently applied to separating the automatic and controlled components of interracial attitudes and behavior on Weapons Identification Tasks (e.g., Amodio, Harmon-Jones, Devine, Curtin, Hartley, & Covert, 2004; Lambert, Payne, Jacoby, Shaffer, Chasteen, & Khan, 2003; Payne, 2001; Payne, Lambert, & Jacoby, 2002; Plant, Peruche, & Butz, 2005).

The primary limitation of this model is that it permits no role for automatic processes that influence behavior even though the correct response can be determined. Thus, this model is not well suited to implicit measures of attitude, such as evaluative priming or the IAT, or for the Stroop Task, in which most people can determine the correct response easily, but an automatic association or habit nevertheless interferes with that response (for a more thorough discussion, see Conrey et al., 2005; Sherman, 2006; Sherman et al., 2008).

As such, when applying this model, researchers should understand clearly exactly what type of automaticity and what type of control the model estimates, and should interpret their results accordingly. The automatic component estimated in this model represents an automatic process that influences responses only when control has already failed (i.e., on a Weapons Identification Task, the automatic component of stereotypes will influence responses only when participants are unable to discern whether a presented object is a gun or a tool). In contrast, most research on automatic attitudes and beliefs seeks to understand the extent to which these constructs are automatically activated in the first place and influence perception and behavior, regardless of the status of control. The controlled component in this model represents control as reflected in the effort expended to accurately identify the nature of a stimulus or situation (e.g., "Is it a gun?"). Thus, control in this model does not represent the type of self-regulatory control that seeks to overcome the influence of unwanted automatic processes.

The Automatic Default model The other original PD model (Lindsay & Jacoby, 1994) is designed to account for tasks in which automatic processes are thought to influence behavior, regardless of whether or not control succeeds. In these tasks, controlled processes drive responses only in the absence of automatic bias. Thus, the model proposes that, in the Stroop Task, if present, an automatic habit to read the word will determine responses. Only in the absence of such a habit will the controlled process of determining the color drive responses. Estimates of the contributions of color perception (control) and reading habit (automaticity) are derived by comparing trials in which the two processes produce the same response (e.g., the word *blue* written in blue ink) with trials in which the two processes would lead to different responses (e.g., the word *blue* written in red ink). This model of process dissociation would appear to be well suited for Stroop-like implicit measures of attitude, such as evaluative priming and the IAT, in which automatic associations influence behavior, regardless of whether or not control succeeds.

The primary limitation of this model is that it does not distinguish between cases in which an automatic association is not activated at all from cases in which the association is activated but is overcome. On the Stroop Task, people provide correct responses on most trials despite the fact that they have an automatic habit to read the word. In these cases, the habit is overcome. In contrast, as described above, a child who cannot read will make few errors simply because he or she has no reading habit to overcome in the first place. This PD model cannot distinguish between these two cases. Likewise, on implicit measures of attitude, the model cannot distinguish between a person who is able to overcome a strong automatic bias and a person who has no bias in the first place (for a more thorough discussion, see Conrey et al., 2005; Sherman et al., 2008).

Again, researchers should be careful to interpret the meaning of the automatic and controlled estimates derived from this model appropriately. In this case, the automatic estimate reflects the extent to which a construct or habit is automatically activated. The controlled estimate reflects the same type of control as in the Control Default model—control exerted to accurately identify a stimulus or situation. Thus, this model also does not estimate control that works to overcome the influence of the automatic process. As described above, this model does not permit automatic biases to be overcome; if the bias is activated, it will drive responses.

The Quad Model

We developed the Quad model (Conrey et al., 2005; Sherman et al., 2008) to address some of the limitations of other PD models, and to provide a more comprehensive analysis of the automatic and controlled components of behavior. The Quad model relies on the method of opposition articulated by the PD approach, and owes a substantial intellectual debt to Jacoby and his colleagues. Like other PD models, the Quad model seeks to separate multiple processing components from performance on a single task. However, the Quad model also differs in important ways from those models.*

Most obviously, whereas original PD analyses produce single estimates of automatic and controlled processing, we believe it is critical to distinguish between two distinct automatic processes and two distinct controlled processes.† In our view, each of these processes represents a ubiquitous and fundamental component of behavior. Indeed, the four processes identified by the Quad model appear repeatedly across a wide spectrum of dual-process models of social psychology (e.g., Chaiken & Trope, 1999; Sherman, in 2008). Of course, by definition, all four processes never appear within any particular dual-process model. Thus, although the shared aim of dual-process models is to assess the extent to which a given judgment reflects relatively automatic or controlled processing, the particular types of automaticity and control of interest vary across models.

Two Types of Control

Dual-process models have generally been concerned with one of two different types of control. In some models, control is characterized by stimulus detection processes that attempt to provide an accurate depiction of the environment. For example, in dual-process models of persuasion, the controlled process discriminates between strong and weak arguments (e.g., Chaiken, 1980; Petty & Cacioppo, 1981; Fazio, 1990). In models of impression formation, the controlled process attends to and integrates target behaviors, providing an individuated (and, presumably, relatively accurate) impression of the person (e.g., Brewer, 1988; Fiske & Neuberg, 1990). This is the type of control represented in Jacoby's PD models.

However, in other dual-process models, control is characterized by self-regulatory processes that attempt to inhibit unwanted or inappropriate information. For example, in Devine's (1989) model of stereotyping, control must be exerted to overcome the automatic influence of stereotypes. In Wegner's (1994) model of thought suppression, control

* See Sherman (2008) for a discussion of the conditions under which two-component PD models and the Quad model may be applied most appropriately.

† Recently, Jacoby and his colleagues (e.g., Jacoby, Bishara, Hessels, & Toth, 2005) have proposed a new PD model that incorporates the two types of automatic processes represented in the Quad model. However, the model does not provide roles for both of the controlled processes estimated by the Quad model and, as a result, cannot separate strength of activation from ability to overcome activation.

must be exerted to inhibit unwanted thoughts. In many models of social judgment, self-regulatory control is exerted when people try to correct their judgments for subjectively expected biases (e.g., Martin, 1986; Wegener & Petty, 1997).

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

Two Types of Automaticity

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

In other dual-process models, however, automatic processes influence behavior only when control fails. Jacoby's (1991) depiction of the role of familiarity in recognition memory is a prominent example of this type of process. Others have portrayed automatic processes in Weapons Identification and Shooter Tasks in this manner (e.g., Amodio et al., 2004; Lambert et al., 2003; Payne, 2001; Payne et al., 2002; Plant et al., 2005). Another example is the implicit preference shown for items on the right side of a display when conscious introspection provides no rational basis for preference (Nisbett & Wilson, 1977).

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

How the Quad Model Works

The Quad model (Conrey et al., 2005; Sherman et al., 2008) is a multinomial model (see Batchelder & Riefer, 1999) designed to estimate the independent contributions of each of the four components described above to a given behavior. More formally, the four components of the model are the automatic activation of an association (*Association ACtivation*, AC), the ability to determine correct and incorrect responses (*Detection*, D), the success at overcoming automatically activated associations, when necessary (*Overcoming Bias*, OB), and the influence of a general response bias that might guide responses in the absence of other available guides to response (*Guessing*, G). Whereas AC and G are automatic processes (though G need not be), D and OB are controlled processes.

The structure of the Quad model is depicted as a processing tree in Figure 13.1. In the tree, each path represents a likelihood. Processing parameters with lines leading to them are conditional upon all preceding parameters. For instance, overcoming bias is conditional upon both association activation and detection. Similarly, guessing is conditional upon the lack of association activation (1-AC) and the lack of detection (1-D). Note that these conditional relationships do not imply a serial order in the onset and conclusion of the different processes. Rather, these relationships are mathematical descriptions of the manner in which the parameters interact to produce behavior. Thus, attempts to detect a correct response and attempts to overcome automatic biases may occur simultaneously. However, in determining a response on a trial of a given task, the influence of attempts to overcome bias will be seen only in cases in which detection is successful.

The conditional relationships described by the model form a system of equations that predict the number of correct and incorrect responses in different conditions. The model's predictions are then compared with actual data to determine the model's ability to account for participants' behavior. A χ^2-estimate is computed for the difference between the predicted and observed errors. In order to best approximate the model to the data, the four parameter values are adjusted through maximum likelihood estimation until they produce a minimum possible value of χ^2. The final parameter values that result from this process are interpreted as relative levels of the four processes. For a complete description of data analysis within the Quad model, see Conrey et al. (2005).

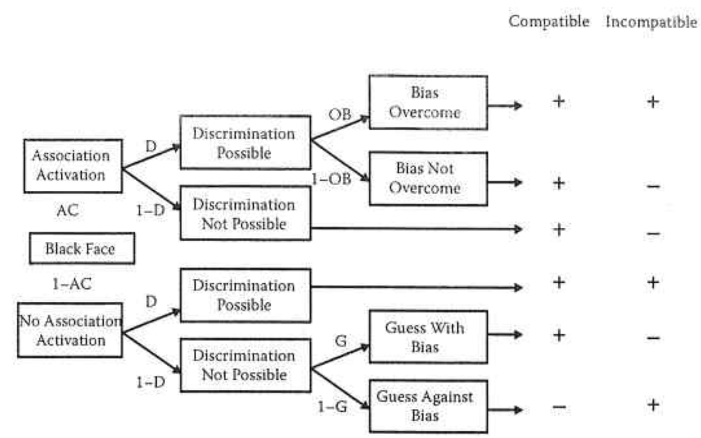

FIGURE 13.1 The Quadruple Process Model (Quad Model). Each path represents a likelihood. Parameters with lines leading to them are conditional upon all preceding parameters. The table on the right side of the figure depicts correct (+) and incorrect (–) responses as a function of process pattern and trial type.

An Example and Range of Application

As an example of how the model works, consider performance on a standard Black-White/positive-negative IAT (Greenwald et al., 1998). The presentation of a Black face may automatically activate negative associations, predisposing the participant to press the *negative* button. Depending on whether the trial is part of a compatible (Black/bad and White/good) or incompatible (Black/good and White/bad) block, this automatic bias will be either congruent or incongruent with the correct answer *Black* achieved through detection. On compatible trials, there is no conflict between what is automatically activated and what is detected. As such, there is no need to overcome the bias in order to produce the correct response. However, on incompatible trials, AC and D generate conflicting responses. Which of these two processes ultimately directs the outcome is determined by whether or not the participant succeeds in overcoming his or her bias. Finally, if no association is activated and the correct response cannot be ascertained, then participants must guess. Guessing need not be random and may, for example, reflect a bias to respond with the positive key when all else fails (Conrey et al., 2005).

The Quad model may be used to analyze data from any measure that compares compatible and incompatible trials, in which automatic and controlled processes are placed in concert with, and in opposition to, one another. Though I have used the IAT as an example, the logic is the same with any implicit measure that compares compatible and incompatible trials, including Stroop Tasks (Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000), evaluative priming tasks (e.g., Fazio et al., 1995), the Weapons Identification Task (e.g., Payne, 2001), the Shooter Task (e.g., Correll, Park, Judd, & Wittenbrink, 2002), and the Go/No-Go Association Task (Nosek & Banaji, 2001). Though the model also may be applied in the same way to explicit measures, we have focused, thus far, on applications to implicit measures. Because of their oft-reputed property of being pure measures of automatic activation, implicit measures are particularly interesting contexts in which to examine the independent contributions of multiple automatic and controlled processes. In contrast, most everyone would probably agree that explicit measures reflect the influence of both automatic and controlled processes.

Results. The viability of a mutinomial model depends on four critical elements: model fit (Can the model adequately approximate the behavioral data?), parameter independence (Can the model's parameters be influenced independently?), construct validity of the parameters (Do the parameters signify the processes claimed by the model?), and predictive validity of the parameters (Do the parameters predict meaningful behaviors?). Without going into great detail, the Quad model has succeeded on all elements (for a full description, see Sherman et al., 2008).

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 (Conrey et al., 2005; Sherman et al., 2008). The parameters also have been shown to vary independently of one another. For example, implementing a response deadline in an IAT reduced D and OB, but left AC and G unaffected. Manipulating the base rate of left-hand versus right-hand responses in the same task affected G, but none of the other three parameters (AC, D, OB). The expectation that one's performance on the weapon identification task (Lambert et al., 2003) would be observed by others decreased participants' ability to accurately detect the stimuli (D), but increased success at overcoming bias (OB). These results indicate that the four parameters of the Quad model can vary independently (for a review, see Sherman et al., 2008).

The construct validity of the model parameters has also been established by a number of findings (Conrey et al., 2005; Sherman et al., 2008). The fact that D and OB were reduced by a response deadline supports the claim that the two parameters reflect controlled processes that require cognitive capacity. In contrast, the finding that AC and 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 decreases with age, two factors 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 IAT (Beer et al., 2008) showed that AC was correlated with activity in the amygdala and insula, which are involved in emotional processing and arousal. This finding is consistent with the depiction of AC as measuring evaluative associations in such a task. At the same time, D was associated with activation in both the dorsal anterior cingulate cortex and the dorsolateral prefrontal cortex, areas of the brain associated with detecting the need for control and implementing control, respectively. This is consistent with the Quad model's depiction of D as a controlled process that selects appropriate behavior and feeds into efforts to overcome inappropriate automatic influences. 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 (see Sherman et al., 2008 for a full review).

Finally, two studies provide evidence for the predictive validity of the parameters. First, estimates of individual subjects' AC parameters derived from an IAT were positively correlated with 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 OB were negatively correlated with reaction time impairment. Thus, the higher the OB, the better able were participants to avoid association-based impairments in performance.

In another study (Gonsalkorale, von Hippel, & Sherman, 2008), a Muslim confederate's ratings of how much he liked interaction partners were predicted by an interaction between AC and OB, derived from participants' GNAT performance. Specifically, when participants had low AC estimates of negative associations with Muslims, their level of OB was unrelated to how much they were liked by the confederate. In contrast, participants with high AC estimates of negative associations with Muslims were liked to the extent that they 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.

Summary

In this section, I have argued that implicit measures should not be taken as direct reflections of the automatic activation of associations in memory. Furthermore, contrasting performance on implicit and explicit measures as a means to estimate automatic and controlled aspects of attitudes is problematic in many ways. Instead, I recommend the use of process dissociation techniques that permit the independent estimation of automatic and controlled components of processing within a single measure of attitude. The Quad model is particularly useful in this regard, as it permits the assessment of multiple automatic and controlled processes that have been shown to be important in many domains of social psychology.

Controlled Influences on Implicit Measures

Perhaps the most controversial claim of the process dissociation approach for implicit attitude researchers is that responses on implicit measures are influenced by controlled processes. Certainly, this view is inconsistent with the common portrayal of implicit measure performance as wholly unintentional, resource-independent, and uncontrollable (e.g., Bargh, 1999; Devine, 1989; Fazio et al., 1995; Greenwald et al., 1998; Kim, 2003). Bargh offered a particularly spirited challenge to the view that implicit attitudes and stereotypes could be controlled. In his entertaining and thought-provoking chapter, he compared these implicit concepts to "cognitive monsters" that social psychologists were desperate to prove could be controlled. In turn, he depicted claims on this controllability as well-meaning but, ultimately, fictional tales. The two central claims of his argument were that (a) the initial activation of implicit attitudes and stereotypes could not be

altered through conscious intent, and that (b), once activated, the use of these concepts could not be inhibited deliberately.

Though the limited data available at the time were vulnerable to skepticism, the intervening years have yielded growing evidence that the fable of the cognitive monster can have a happier ending than Bargh (1999) deemed likely. Still, under careful scrutiny, some aspects of the case for controllability remain fragile, and the cognitive monster is not yet completely subdued. An examination of this literature through the lens of the Quad model helps to clarify which aspects of controllability do and do not have strong empirical support. First, I will discuss conceptual reasons to believe that the extent of controllability has been underestimated. I also will discuss problems in distinguishing between controlling the initial activation versus application of automatic associations. Then, I will review the empirical evidence for controllability. In so doing, I will consider both indirect evidence gleaned from studies on the malleability of implicit attitudes as well as more direct evidence of controllability.

Reasons for the Underestimation of Controllability

For a number of reasons, the extent of control over implicit attitudes has probably been underestimated. Most basically, there is the almost universal view that, by definition, responses on implicit measures cannot be controlled. Given this stance, it is not surprising that researchers have been reluctant to attribute fluctuations in implicit attitudes to the influence of controlled processes. Again, this view may be contrasted with that of the process dissociation approach, which assumes that all measures, including implicit ones, are influenced by both automatic and controlled processes.

Another factor in the underestimation of controllability is the standard treatment of stimulus onset asynchronies (SOAs) and reaction-time data. In priming studies, the SOA is the amount of time that elapses between the onset of a prime and the appearance of the target. Neely (1977) first reported that at SOAs shorter than 500 ms, priming effects occurred, regardless of participants' intentions to respond otherwise. This formed the basis for what has since become the standard view that control is not possible with short SOAs. However, there are two significant problems with this assumption. First, it is not only the SOA, but also the time taken to respond to the target that influences the extent of controllability. This is easily demonstrated with a simple thought experiment. Imagine that a face is presented as a prime for 200 ms and, after a brief pause of 200 ms, a positive word is presented, to which participants must respond by pressing a positive key on their computers. That would be a 400-ms SOA and, in the eyes of many, would, therefore, be uncontrollable. But imagine that a participant sits there looking at the positive word, thinking about her response for 1, 2, 3, 10 seconds. After 10 seconds, she finally presses the key. Is this an automatic response? Certainly not. Thus, a brief SOA does not guarantee that a participant will not sit there consciously pondering the nature of her response long past the time at which it could reasonably be considered to be "automatic."

There are many findings in the implicit attitudes and stereotyping literature that are interpreted as reflecting automatic bias on the basis of brief SOAs, even though the subsequent target reaction times may be as long as 1000 ms or more. A 500-ms SOA plus a 1000-ms reaction time represents a response that has taken place a full second and a half after the initial presentation of the prime. Even an SOA of 200 ms plus a 500-ms reaction time likely leaves more room for control than has been acknowledged. The point is that a brief SOA is no guarantee of uncontrollability, and we do not have careful metrics of what SOA/reaction-time combinations do and do not permit control on different tasks. This lack of neat boundaries between automatic and controlled SOAs/reaction times is yet another incentive to apply process dissociation techniques to identify the automatic and controlled components of implicit task performance.

The second problem with this treatment of SOAs and reaction times is that there are now a number of documented cases where control appears to be present at very brief SOAs (e.g., Blair & Banaji, 1996; Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002; Glaser & Banaji, 1999; Kawakami et al., 2000; Maddux, Barden, Brewer, & Petty, 2005; Moskowitz, Gollwitzer, Wasel, & Schall, 1999). There are two possible interpretations of such findings. One is that participants engage in *automatic control* or *automatic inhibition*. Though inhibition can surely be routinized to the point of automaticity, this explanation raises the specter of circularity. That is, if control is, by definition, not possible at brief SOAs, then these effects must be attributed to automated control. The second interpretation—that participants are actually able to apply control in these situations—has rarely been invoked.

Another reason for the underestimation of control concerns the role of awareness in implementing controlled processes. It is generally assumed that biases may be controlled only if a person is consciously aware both that control is necessary and of how to correct for the bias (e.g., Bargh, 1999; Strack, 1992; Tesser & Martin, 1996; Wegener & Petty, 1997). The lack of either or both kinds of awareness is one of the central bases for the designation of implicit measures as implicit (e.g., Fazio et al., 1995; Greenwald et al., 1998; see this volume) and one of the key sources for the conclusion that responses on these measures cannot be controlled.* However, the process dissociation perspective views the role of awareness differently. For example, both of Jacoby's models and the Quad model propose that one type of controlled process, stimulus detection, plays a role in all responses on implicit measures, regardless of whether participants are intentionally trying to correct a perceived bias. Factors that influence these detection processes, such as a speeded judgment task or an enhanced motive to perform a task accurately, affect implicit task performance even if participants are not aware of implicit bias or how to control it. There are now many process dissociation results demonstrating such effects on controlled components of implicit responses (see below; Conrey et al., 2005; Lambert et al., 2003; Payne, 2001; Payne et al., 2002; Plant et al., 2005).

The Quad model proposes that overcoming the influence of automatic associations is an additional controlled process that regularly occurs during implicit task performance. This process does not depend on conscious awareness that overcoming bias is necessary or on applying a conscious strategy to overcome such bias. Rather, overcoming bias may occur whenever there is a conflict between responses suggested by automatic associations and stimulus detection, even if there is no awareness of the bias or of a strategy for countering it. Thus, on a priming task, a Black face may activate negative content and push responses in that direction. When a positive target word appears, participants must overcome the influence of the prime to respond correctly to the target. However, they need not be consciously aware that the prime is affecting their responses or that the prime is related in any way to the target. Indeed, overcoming the influence of an entirely subliminal prime should proceed in much the same way. All participants need to know is that they would like to provide a particular response on the task. The fact that that response happens to be incompatible with a prime engages overcoming bias processes. These processes are controlled in that they require time and resources, and may be influenced by intention (Conrey et al., 2005). Thus, from the process dissociation perspective, the

* Some researchers have argued that people generally are aware of their implicit biases (e.g., Fazio & Olson, 2003; Monteith et al., 2001). However, even when this is true, people are not typically aware of how that bias influences responses on implicit measures and cannot formulate conscious strategies to counteract it (e.g., Kim, 2003).

stipulation that control requires awareness has led to an underestimation of the extent of control.

Controlled Activation Versus Application

A key distinction in considering the controllability question is the extent to which control is exerted over the initial automatic activation of underlying associations versus the application of already-activated associations (e.g., Bargh; 1999; Brauer, Wasel, & Niedenthal, 2000). The common view that implicit measures provide a direct, uncontaminated picture of the information that is automatically activated in memory complicates assessments of this distinction. Namely, this view dictates that any evidence of controlled influences on implicit task performance must be taken as a reflection that the initial activation of associations can be controlled (e.g., Blair, Ma, & Lenton, 2001; Govan & Williams, 2004; Kawakami et al., 2000; Lepore & Brown, 1997; Maddux et al., 2005; Moskowitz et al., 1999; Lowery, Hardin, & Sinclair, 2001). The possibility that variations in implicit task performance might be due to controlled processes that alter the expression of the same activated content is, generally, not considered. As a result, research on implicit attitudes has likely overestimated the ease with which people can intentionally alter their underlying associations and underestimated the extent to which these automatically activated associations can be controlled, once activated.

Viewed from the perspective of the Quad model, demonstrations of controlled influences on implicit measures may reflect control of activation, application, both, or neither. Specifically, according to the Quad model, attempts at control may change the underlying associations that are automatically activated, the ability to accurately detect and represent the environment, the ability to overcome activated associations, or the nature of response biases. In the case of D and G, implicit task performance may be altered even if there are no associations automatically activated in the first place (see Figure 13.1). Of course, one of the most significant advantages of the Quad model is that these different effects of control need not be exclusive. All four components may be influenced by attempted control, and the Quad model is able to identify each of these independent effects from performance on a single implicit measure.

Evidence for Controllability: Implicit Attitude Malleability

The preceding discussions suggest that control influences implicit attitudes to a greater extent than has been recognized. What evidence is

there to support this claim? One potential source is the now extensive body of research showing that various experimental, contextual, motivational, and personal factors can alter the nature of responses on implicit measures. Indeed, these findings have been regarded as a challenge to the view that implicit attitudes cannot be controlled (e.g., Blair, 2002). However, there are two key questions to consider in evaluating these findings. First, when such effects are observed, does control play any role at all? There is an important distinction between showing that implicit task performance is malleable and showing that it is controllable. The fact that the identical implicit attitude is not inevitably produced on a measure, regardless of variations in stimuli, context, motives, or individual differences, does not necessarily invoke the presence of control. For example, these variations may passively alter the accessibility of different contents in memory, regardless of any exertion of intent or control. Second, if control is involved, at what stage is it exerted? Is it exerted over the initial automatic activation of associations or is it exerted over the application of those associations in responding to the implicit task demands?

The extent to which the malleability findings answer these questions is obscured by the common position that, by definition, implicit measures reflect only the automatic activation of content from memory, and they cannot be controlled. As detailed above, this view has likely led to a general underestimation of the role of control and an overemphasis on the control of activation versus application of automatic associations. Deciphering the role of control in these findings also is difficult because the same behavioral outcome may be explained by a variety of different underlying processes, some of which are controlled, and some of which are automatic. For example, fluctuations in implicit attitudes may reflect fluctuations in any of the four underlying processes in the Quad model. A major advantage of the process dissociation approach is that it can separate these components and specify their independent contributions. However, with only the behavioral data to consider, what follows necessarily will be largely speculative.

Stimulus feature effects. In some cases, controlled processes would seem to play a modest role, at best. For example, contextual variations in stimulus features (e.g., Barden, Maddux, Petty, & Brewer, 2004; Wittenbrink, Judd, & Park, 1997) may influence performance on implicit tasks by altering the associations that are automatically activated without affecting controlled processes at all. In this case, variations in implicit task performance would reflect the simple fact that different stimuli activate different associations in memory. As such, these data may not imply that a given attitude is either malleable (i.e., the data do not show that a given attitude is changed but that different attitudes are activated) or controllable. Such demonstrations of contextual variation in implicit activations are certainly important for understanding the nature of category representation and the specificity of automatic processes. However, it is not clear that they have much to do with the debate on controllability.

Accessibility and context effects. Controlled processes may play a larger role in other demonstrations of fluctuations in implicit measures. For example, variations in accessible category exemplars (Dasgupta & Greenwald, 2001; Govan & Williams, 2004), accessible category features (e.g., Blair et al., 2001; Livingston & Brewer, 2002), situational contexts (Lowery et al., 2001; Richeson & Ambady, 2003; Richeson & Nussbaum, 2004; Rudman, Ashmore, & Gary, 2001; Sechrist & Stangor, 2001), or situationally induced motives (e.g., Sinclair & Kunda, 1999) may influence implicit measures by enhancing efforts at controlling what associations are activated in memory, efforts to perform the task more accurately, efforts to overcome associations, or guessing biases. However, it also is possible that these effects are driven entirely by passive changes in the associations that are activated by the different conditions, without the intervention of control at all. If so, then these effects would, again, represent the contextual specificity of association activation rather than the malleability or controllability of a given attitude. Indeed, these findings have most commonly been interpreted as reflecting unintended changes in activation (e.g., Dasgupta & Greenwald; Govan & Williams; Livingston & Brewer; Lowery et al.; Rudman et al.). As described below, we have preliminary evidence that this is, in fact, the case (Sherman et al., 2008).

Individual differences. Another potential source of evidence for controllability are the numerous studies demonstrating individual differences in implicit attitudes (e.g., Devine et al., 2002; Maddux et al., 2005; Moskowitz et al, 1999). The favored explanation for such differences is that people who are properly motivated and, therefore, practiced at overcoming unwanted biases may automate the process to the extent that they can control their responses even on implicit measures. However, though the development of such expertise is clearly one way to gain control over implicit biases, interpreting such individual differences can be difficult. In particular, it often is difficult to rule out the possibility that people who show less bias on implicit measures do so, not because they are efficient at regulation, but because they simply have weaker or different associations to regulate in the first place (e.g., Conrey et al., 2005; Lepore & Brown, 1997; Sherman et al., 2008). Even person X situation interactions may be interpreted as showing that in different contexts, different associations are accessible for different people. Simply put, it is difficult to distinguish differential activation from differential inhibition (for other examples, see Sinclair & Kunda, 1999; Kawakami et al., 2000). This suggests that the role of controlled processes in these individual differences may be smaller than has been proposed.

At the same time, other considerations and findings suggest that the role of control may have been underestimated in explaining these individual differences. First, as described above, what is typically described as automatic control may, in fact, be standard control (e.g., is subject to intention, resources, etc.). The designation of this control as automatic is based almost entirely on the a priori assumption that control is simply not possible on implicit measures. From the perspective of the Quad model, an increased ability among some individuals to overcome bias need not represent an automatic process (e.g., automatic inhibition), per se. That is, skill at overcoming bias may be enhanced even though the process retains features of a controlled process (e.g., influenced by intent, resource availability, etc.).

Research by Amodio and his colleagues indicates that individual differences in implicit biases may have more to do with controlled conflict detection processes than with either automatic activation or inhibition processes (see below; Amodio, Devine, & Harmon-Jones, 2008; Amodio et al., 2004; Sherman et al., 2008). In particular, these studies have provided evidence that what sets apart people who show weaker or no implicit bias is neither the strength of their automatic associations (but see Sherman et al., 2008) nor their ability to overcome those associations. Rather, their particular skill appears to lie in their increased sensitivity to detecting conflict among behavioral impulses (e.g., conflicts between automatically activated associations and controlled determinations of accurate responses). Detecting such conflict is itself a controlled process, and superior detection skills may facilitate the application of other self-regulatory controlled processes that help to overcome intrusive biases that thwart desired (or task-accurate) responses (see below; Monteith, Ashburn-Nardo, Voils, & Czopp, 2002).

Summary. In summary, these demonstrations of implicit attitude malleability certainly refute any claim that the exact same implicit attitude will be educed inevitably by all stimuli, for all people, and in all contexts. Much less clear is the extent to which these findings represent a significant challenge to the claim of Bargh (1999) and others that implicit attitudes cannot be controlled. Due, in part, to the standard conceptions of implicit measures, few researchers have interpreted their results as demonstrating controlled processes. Instead, the results primarily have been interpreted as reflecting either changes in the content that is automatically activated or the application of automatized control. Initial attempts to address these questions with the Quad model suggest that both automatic and controlled processes are involved in these effects.

One way to better understand the bases of these different malleability effects is through the application of process dissociation techniques. In our own research, we have attempted to do so by examining changes in the Quad model parameters that correspond to changes in implicit task performance (i.e., implicit attitudes). Thus, we are directly measuring the extent to which a change in the behavioral measure is related to changes in the activation of different associations, changes in participants' ability to detect correct and incorrect responses, changes in overcoming associations, or changes in response biases.

As one example, we have applied the Quad model to examine the effects of altering the accessibility of category exemplars on implicit bias (Sherman et al., 2008). Replicating past research, participants exposed to positive Black and negative White exemplars showed weaker implicit pro-White bias than participants who were shown positive White and negative Black exemplars. Our modeling results suggest that this effect is because different associations are activated in the pro-Black condition than in the pro-White condition. Thus, exposure to different exemplars alters the nature of the associations that are automatically brought to mind in performing the IAT, but not participants' ability to detect correct responses or overcome bias. Other applications of the Quad model to questions of malleability are ongoing, including research aimed at understanding the bases of individual differences in implicit attitudes (Sherman et al., 2008).

Direct Evidence of Control

Process Dissociation Findings

More direct evidence of controlled influences on implicit attitudes comes from a variety of sources. First, there is the accumulating body of evidence from process dissociation analyses. Researchers applying Jacoby's models have reported a number of contributions of the controlled component to responses on implicit measures. The tasks that have produced such effects include Payne's Weapons Identification Task (e.g., Amodio et al., 2004; Lambert et al., 2003; Payne, 2001; Payne et al., 2002) and the Shooter Task (Plant et al., 2005). These findings have shown not only that PD can be applied to such tasks but also that the controlled component of PD within these tasks responds in predictable ways. For example, factors that diminish the opportunity for controlled processing, such as rapid responding (e.g., Payne, 2001; Payne et al., 2002) or social distraction (e.g., Lambert et al., 2003), have been shown to diminish the PD estimates of control, whereas factors that should enhance control, such as extensive practice (e.g., Plant et al., 2005), have been shown to increase the PD estimate of control.

Application of the Quad model (Conrey et al., 2005) has further shown that multiple controlled processes contribute to implicit task performance. Both stimulus detection and overcoming bias have been shown to play a role in the IAT, the GNAT, the WIT, and other priming tasks (Conrey et al., 2005; Sherman et al., 2008). These studies also have shown that the controlled components of the Quad model respond to manipulations in predicted ways. Thus, both D and OB are reduced by a short response window. In contrast, the expectation of an audience diminishes D but enhances OB. Finally, OB is reduced both by aging and alcohol consumption.

Detection and Overcoming Bias also have demonstrated meaningful predictive validity. In the IAT, the extent of OB was negatively correlated with the strength of the reaction time bias, suggesting the role of overcoming automatic associations in the extent of that bias. In our Muslim interaction study (Gonsalkorale et al., 2008), OB (from a GNAT) interacted with automatic associations to influence a confederate's impressions of participants. When participants had weak associations, OB was unrelated to the confederate's impressions. In contrast, participants with strongly biased associations were liked to the extent that they had high OB parameter estimates. Thus, the ability to overcome automatic negative associations predicted impressions when those associations were strong.

The role of D in reducing implicit prejudice has been demonstrated recently by examining a group of people that Devine and her colleagues have identified as better able to respond without bias on implicit measures of prejudice (Devine et al., 2002). These "good regulators" have been distinguished from "poor regulators" (who wish to be non-prejudiced, but fail to achieve that goal on implicit measures) and "non-regulators" (who are not concerned about responding in prejudiced ways). Good regulators have been defined as those who are high on Internal Motivation to Respond Without Prejudice (IMS) but low on External

Motivation to Respond Without Prejudice (EMS; see Devine et al., 2002 for further details). Poor regulators are high on both IMS and EMS. Non-regulators are those who are low on IMS and, thus, have no internal motivation to be non-prejudiced. What accounts for the differences among these groups? One possibility is that they differ in the nature of their automatic associations (AC), with good regulators possessing less biased associations than the other groups. Another possibility is that good regulators are better at determining correct and incorrect responses (D) than the other groups. Finally, a third possibility is that good regulators, as their name would suggest, are better at overcoming their biases (OB) than the other groups. The Quad model provides a means for distinguishing among these accounts, and we (Sherman et al., 2008) applied the model to data collected from a WIT (Amodio et al., 2008). There were two major findings. First, we found that both good and poor regulators had less biased automatic associations than non-regulators. Second, we found that the key factor separating the good and poor regulators was the detection parameter (D), and not the underlying associations (AC) or the ability to overcome them (OB).

Further Evidence for an Implicit Task Cognitive Skill

Other evidence also indicates that general cognitive abilities may underlie performance on implicit measures. In particular, McFarland and Crouch (2002) reported high within-participant correlations on a wide range of IATs. Thus, participants who showed strong or weak bias on one kind of IAT (e.g., having to do with race) tended to show a similar strength of bias on other IATs (e.g., having to do with taste).* These data cannot be explained in terms of the strength of the underlying associations, but rather must be due to differing levels of skill at meeting the task demands of the IAT (e.g., resolving incompatible influences on response). It also seems unlikely that these skilled individuals have automatized control in each of the different content domains tested in the research. Instead, it would appear that some individuals are simply better able to implement control in performing the task.

* Though the extent of these correlations was reduced by the application of a different scoring procedure, they were not eliminated (Cai, Sriram, Greenwald, & McFarland, 2004).

Intentional Regulation

Other research has demonstrated that conscious intentions to avoid bias can be effective. Kawakami et al. (2000) showed that, after hundreds of trials of intentionally negating stereotypes, participants showed weaker automatic stereotyping effects. We (Sherman et al., 2008) performed a similar training study and examined the effects on IAT performance. Replicating Kawakami et al. (2000), we found that negation training reduced implicit bias. Application of the Quad model to the data showed that the training not only weakened participants' automatically activated associations (AC), but also improved their ability to determine the correct response (D).

Neuroscientific evidence of control. Finally, there also is growing neuroscientific evidence for the role of control in implicit attitudes. Using fMRI, Chee and his colleagues (Chee, Sriram, Soon, & Lee, 2000; see also Luo et al., 2006) showed that performing an insect-flower IAT involves the left dorsolateral prefrontal cortex (DLPFC) and the anterior cingulate cortex (ACC). Whereas the DLPFC is associated with inhibitory processes, the ACC is associated with conflict detection and resolution processes. Both types of processes are considered to be executive, controlled operations.

Using fMRI, Richeson and her colleagues (Richeson, Baird, Gordon, Heatherton, Wyland, Trawalter, & Shelton, 2003) showed the involvement of the DLPFC when White participants were simply exposed to Black faces. This suggests that mere exposure to targets of potential bias is sufficient to instigate controlled processes. Moreover, the extent of DLPFC activation was related to the strength of implicit prejudice. Thus, participants with the strongest implicit prejudice apparently worked the hardest to control their behavior. These efforts were shown to deplete regulatory resources on a subsequent Stroop Task.

Also using fMRI, Cunningham and his colleagues (Cunningham, Johnson, Raye, Gatenby, Gore, & Banaji, 2004) reported the involvement of the PFC and the ACC when White participants were exposed to Black and White faces. Moreover, activation in these regions modulated the extent of amygdala activation when the faces were presented for 525 ms (vs. 30 ms). The amygdala is associated with automatic emotional reactions and has been shown to correlate with implicit racial bias (see also Phelps et al., 2000). This suggests that the PFC and ACC were actively regulating affective responses produced by exposure to Black and White faces, at least when there was sufficient time for control to engage.

Finally, using both a different behavioral measure and a different neurocognitive measure, Amodio and his colleagues (Amodio et al., 2004) also found evidence that the ACC is active during implicit task performance. In particular, Amodio et al. used electroencephalographic (EEG) recording to measure brain activity as participants performed a Weapons Identification Task (Payne, 2001). Their analysis focused specifically on the extent of error-related negativity (or ERN) during the task. The ERN is sensitive to response conflicts that result in task errors, and has been shown to originate from neural activity in the ACC (Dehaene, Posner, & Tucker, 1994) and to predict subsequent behavioral control (e.g., Gehring, Goss, Coles, Meyer, & Donchin, 1993). The results showed enhanced ERN activity during the commission of race-based errors on incompatible trials (i.e., Black faces priming pictures of tools). Amodio et al. also performed a process dissociation analysis and found that the extent of race-bias ERNs was associated with higher estimates of the controlled component of PD. These results strongly suggest the involvement of controlled processes in the execution of the priming task (see also Sherman et al., 2008).

Summary. In summary, there is growing direct evidence that controlled processes do influence responses on implicit measures. The process dissociation findings show that controlled processes are a fundamental component of implicit task performance. Moreover, the extent of this control is influenced by motivated intentions. Specifically, the likelihood of overcoming automatic associations is enhanced by efforts to appear nonprejudiced in a social context (Conrey et al., 2005). Also, the ability to detect appropriate behavior is related to internal motivations to act in nonprejudiced ways (Amodio et al., 2008; Sherman et al., 2008) and reduces implicit prejudice. Obversely, those who lack such detection skills demonstrate implicit prejudice even though they would prefer not to. The work of McFarland and Crouch (2002) similarly implicates the role of controlled skills in performing implicit tasks. There is also evidence that conscious training may facilitate this process (e.g., Kawakami et al., 2000; Sherman et al., 2008). Finally, there is now considerable neuroscientific evidence that brain regions associated with control are actively engaged upon exposure to bias-prone targets and during implicit task performance (Amodio et al., 2004; Chee et al., 2000; Luo et al., in press; Richeson et al., 2003), and that the activity of these regions modulates automatic affective responses (Cunningham et al., 2004).

These data provide evidence of control over both the initial activation of automatic associations and their subsequent application. The Quad model analyses reported by Sherman et al. (2008) showed that both good and poor regulators had less biased associations than non-regulators. These different associations could have resulted from motives to be non-prejudiced or through long-term practice aimed at altering those associations. However, as with the other individual difference data (e.g., Devine et al., 2002; Maddux et al., 2005; Moskowitz et al, 1999), it is impossible to rule out the possibility that the good and poor regulators simply never had strongly biased associations. Relatedly, and perhaps more compelling, the Quad model analysis of our negation study (Sherman et al., 2008) showed that negation training reduced the activation of biased associations in a subsequent IAT. Whether the training led to "automatic inhibition" of the associations, altered those associations, or instead activated alternate associations is unclear. However, in any case, the intentional training did alter the associations that were automatically activated.

Most of the other findings suggest control over the application of activated associations. For example, the skills of good regulators (i.e., high IMS/low EMS ability to detect appropriate behavior) identified by Sherman et al. (2008) and the general skills identified by McFarland (McFarland & Crouch, 2002) appear to correspond to controlled processes that exert influence during implicit task performance. Other research showed that impairments in the ability to control the application of activated associations were associated with increased implicit bias. For example, increased implicit bias in public (vs. private) settings seems to be related, at least in part, to diminished ability to determine correct behavior in such contexts (Conrey et al., 2005; Lambert et al., 2003). In contrast, the effects of aging and alcohol consumption on implicit bias were shown to be related to failures in the ability to overcome activated associations (Sherman et al., 2008). Diminished ability to overcome activated associations was also shown to be related to increased reaction time bias on the IAT and to making poor impressions during an interaction with a Muslim confederate (Sherman et al., 2008). Finally, the brain regions shown to be active during the performance of implicit tasks and during exposure to race-prone targets are those that influence behavior post-activation. Specifically, they have to do with detecting conflict produced by opposing activations and resolving that conflict with self-regulatory control (e.g., Amodio et al., 2004; Chee et al., 2002; Cunningham et al., 2004; Richeson et al., 2003). Other research has also shown that people can control the application of stereotypes that have been activated via thought suppression without awareness or intent (e.g., Wyer, Sherman, & Stroessner, 2000, Experiment 2). In sum, this research indicates that the ability to control the influence of automatic associations once they have been activated has been significantly underestimated and that, perhaps, the ability to control the nature of the associations that are activated has been overestimated.

Conclusion

The advent of implicit measures has revolutionized research on attitudes by offering a solution to the perpetual problems of willing and able. Though the advances produced by this approach are undeniable, there are significant drawbacks with the manner in which these tasks are commonly used and interpreted. In particular, the use of task dissociation paradigms to estimate the automatic and controlled components of attitudes introduces a number of significant methodological and conceptual problems. In this chapter, I argue that there are many advantages to a process dissociation approach that provides independent estimates of the simultaneous contributions of automatic and controlled processes to a single task. The Quad model (Conrey et al., 2005; Sherman et al., 2008) provides a nuanced depiction of the operation of multiple, distinct automatic and controlled contributions to behavior.

I also argue that, for a variety of reasons, the extent to which implicit attitudes are subject to control has been underestimated substantially. The process dissociation approach provides a very different perspective than the standard view on the role of control in implicit attitudes and measures, and provides means of estimating the extent of this control. The Quad model, in particular, also permits a much clearer distinction between controlling the automatic activation of content from memory and controlling the subsequent application of that content in performing a given task. From this view, it appears that the extent to which people are able to control the initial activation of automatic associations has been overestimated and that the ability to control the expression of those associations (including on the implicit tasks themselves) has been underestimated. This unacknowledged role for control also has likely concealed the true extent of automatic bias. Thus, the cognitive monster of automatic associations (Bargh, 1999) may be controlled once it has been summoned, but altering the nature of the beast that arrives may be very difficult.

References

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

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

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.

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

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

Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70*, 1142–1163.

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.

Brauer, M., Wasel, W., & Niedenthal, P. (2000). Implicit and explicit components of prejudice. *Review of General Psychology, 4*, 79–101.

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.

Cai, H., Sriram, N., Greenwald, A. G., & McFarland, S. G. (2004). The Implicit Association Test's D measure can minimize a cognitive skill confound: Comment on McFarland and Crouch (2002). *Social Cognition, 22*, 673–684.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*, 752–766.

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

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

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.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of Black and White faces. *Psychological Science, 15*, 806–813.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 121*, 163–170.

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.

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science, 5*, 303–305.

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.

Dovidio, J. F., & Fazio, R. H. (1992). New technologies for the direct and indirect assessment of attitudes. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 204–237). New York: Russell Sage Foundation.

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

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 unobstrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013–1027.

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

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.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*, 385–390.

Glaser, J., & Banaji, M. R. (1999). When fair is foul and foul is fair: Reverse priming in automatic evaluation. *Journal of Personality and Social Psychology, 77*, 669–687.

Gonsalkorale, K., von Hippel, W., Sherman, J. W., & Klauer, K. C. (2008). Bias and regulation of bias in intergroup interactions: Implicit attitudes toward Muslims and interaction quality. Unpublished manuscript.

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

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

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., 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., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 122*, 139–154.

Kawakami, K., & Dovidio, J. F. (2001). The reliability of implicit stereotyping. *Personality and Social Psychology Bulletin, 27*, 212–225.

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.

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.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology, 72*, 275–287.

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.

Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology, 82*, 5–18.

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.

Luo, Q., Nakic, M., Wheatley, T., Richell, R., Martin, A., & Blair, R. J. R. (2006). The neural basis of implicit moral attitude—An IAT study using event-related fMRI. *NeuroImage, 30*, 1449–57.

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.

Martin, L. L. (1986). Set/reset: Use and disuse of concepts in impression formation. *Journal of Personality and Social Psychology, 51*, 493–504.

McFarland, S. G., & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition, 20*, 483–510.

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.

Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition, 19*, 395–417.

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.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226–254.

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.

Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science, 14*, 636–639.

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., 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.

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

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.

Plant, E. A., Peruche, B. M., & Butz, D. A. (2005). Eliminating automatic racial bias: Making race non-diagnostic for responses to criminal suspects. *Journal of Experimental Social Psychology, 41*, 141–156.

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

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

Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology, 40*, 417–423.

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45*, 9, 1043–1056.

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology, 81*, 856–868.

Sechrist, G. B., & Stangor, C. (2001). Perceived consensus influences intergroup behavior and stereotype accessibility. *Journal of Personality and Social Psychology, 80*, 645–654.

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review, 115*, 314–335.

Sherman, J. W. (2006a). Clearing up some misconceptions about the Quad Model. *Psychological Inquiry, 17*, 269–276.

Sherman, J. W. (2006b.0). On building a better process model: It's not only how many, but which ones and by which means. *Psychological Inquiry, 17*, 173–184.

Sherman, J. W., Gawronski, B., Conrey, F. R., Hugenberg, K., & Groom, C. (2006). *The Quad Model of impulse and self-regulation*. Unpublished manuscript.

Sinclair, L., & Kunda, Z. (1999). Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77,* 885–904.

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

Steffens, M. C. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology, 5,* 165–179.

Strack, F. (1992). The different routes to social judgments: Experiential versus informational strategies. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments.* Hillsdale, NJ: Erlbaum.

Stroop, J. R. (1935). Studies on the interference in serial verbal reactions. *Journal of Experimental Psychology, 59,* 239–245.

Tesser, A., & Martin, L. (1996). The psychology of evaluation. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 400–432). New York: Guilford Press.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. *Advances in Experimental Social Psychology, 29,* 141–208.

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

Wilson, T. D., & Hodges, S. D. (1992). Attitudes as temporary constructions. In L. Martin and A. Tesser (Eds.), *The construction of social judgment* (pp. 37–65). Hillsdale, NJ: Erlbaum.

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.

Wyer, N. A., Sherman, J. W., & Stroessner, S. J. (2000). The roles of motivation and ability in controlling the consequences of stereotype suppression. *Personality and Social Psychology Bulletin, 26,* 13–25.

Section VII

Additional Measures