Title

Mathematical modeling of implicit social cognition: The machine in the ghost

Authors
Sherman, JW
Klauer, KC
Allen, TJ

Publication Date
2023-12-10

Peer reviewed

CHAPTER 9

Mathematical Modeling of Implicit Social Cognition

The Machine in the Ghost

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

Although implicit measures often are portrayed as pure indices of automatic associations, instead they reflect the joint contributions of a variety of processes that may be automatic or controlled component processes. Mathematical modeling of implicit task performance attempts to identify and measure these processes. This approach can shed light on a number of important theoretical and empirical issues surrounding implicit social cognition. Modeling also avoids some significant methodological problems inherent in other means of measuring automatic and controlled aspects of social cognition. The first part of this chapter describes some difficulties in the interpretation and use of implicit measures and the related advantages of modeling approaches. The second part of the chapter reviews specific models that have been used to account for performance on various implicit measures and the questions to which they have been applied.

However, before turning to the central concerns of this chapter, some definitional issues require comment. In discussing implicit social cognition, it is important to distinguish among features of the measurement procedure, the behavioral responses obtained with those procedures, and the psychological constructs those responses are meant to reflect. In this chapter, we refer to measurement procedures that assess attitudes and knowledge indirectly (i.e., without explicitly asking people to report their attitudes and knowledge) as implicit measures or implicit tasks. The term *indirect measure* may be technically more accurate for our intended meaning (see De Houwer & Moors, Chapter 10, this volume), but we nevertheless use the common terminology of *implicit*.

Behavioral responses on these measures (e.g., button presses measured in reaction times or error rates) are referred to as behavioral responses or behavioral bias. In the implicit social cognition literature, these behavioral responses usually are labeled as implicit attitudes or implicit knowledge (e.g., stereotypes). We, too, use these terms when it would be awkward, abnormal, or confusing to do otherwise. However, their use may be a source of misunderstanding. In some cases, these terms are meant to imply a specific representational model, most typically that the behavioral responses reflect underlying associations in memory (the presumed underlying psychological construct). Yet in other cases, the labels are used simply as a way to describe the behavioral outcomes of implicit measures and do not carry strong implications about mental representation. One central aim of the modeling approach is to specify the extent to which behavioral responses reflect the activation of associations versus other response processes. As such, it would make little sense to promote the assumption that the behavioral responses reflect underlying associations. Thus, in this chapter, use of the term *implicit attitude* or *implicit stereotype* carries no representational assumption.

Because for purposes of this chapter *implicit attitude* simply refers to an attitude that is measured with an implicit measure, our use of the term also implies nothing about the automatic nature of the hypothetical constructs (e.g., evaluative associations) and processes (e.g., detection of correct responses) that generate behavioral response biases. Those underlying constructs and processes may or may not influence responses efficiently (e.g., in the absence of time or processing resources), without awareness or without intention, or be difficult to inhibit (Bargh, 1994). Which, if any, of these varying components of automaticity apply to a particular construct or process can be determined only through empirical research. The components do not necessarily covary with one another, and a given construct or process may possess features of both automaticity and control. Furthermore, these features may change over time (e.g., through extensive practice, an initially controlled process may acquire features of automaticity). Detailed analyses of the precise manners in which given constructs or processes are automatic is beyond the scope of this chapter (see Moors, Spruyt, & De Houwer, Chapter 2, this volume).

A BRIEF OVERVIEW OF MATHEMATICAL MODELING

This chapter is not meant to provide a comprehensive description of mathematical modeling approaches in psychology. Such accounts can be found elsewhere (e.g., Luce, 1995, 1999; Myung & Pitt, 2002). Nevertheless, before we turn to the core of this chapter, a very brief sketch of the nature of mathematical modeling may be useful for readers unfamiliar with the approach.

Mathematical modeling of implicit measures seeks to identify and quantify the processes that account for performance on the measures. To do so, models attempt to describe outcomes on the measures (error rates, reaction times) via a set of variables (or parameters) and a set of equations that establish relationships among the variables. The variables in the equations represent the hypothesized component processes (e.g., activation of associations, detecting a correct response, overcoming bias, response bias). Solving for these variables yields estimates of the extent of the processes. In some cases, such as with Signal Detection Theory (SDT; Correll, Park, Judd, & Wittenbrink, 2002; Green & Swets, 1966) or Process Dissociation (PD; Jacoby, 1991; Payne, 2001), the equations can be solved algebraically. In other cases, such as with multinomial models (e.g., Batchelder & Riefer, 1999; Sherman et al., 2008; Stahl & Degner, 2007) or diffusion models (e.g., Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007; Ratcliff, 1978), parameter estimates are systematically varied through maximum-likelihood estimation or related procedures to determine the values that most closely reproduce actual task performance.

There are two main purposes of modeling. First, it is used to identify the processes that best account for performance on the task of interest (i.e., model fitting). Models can be compared on this dimension to determine their relative merits in describing a task. Second, modeling is used to extract measures of component processes that may then function as distinct variables. For example, the Quad model (e.g., Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Sherman et al., 2008) estimates four qualitatively distinct processes: Association Activation, Detection, Overcoming Bias, and Guessing. The component process estimates provided by a model may be used to predict individual differences in traits or motives (e.g., Amodio, Devine, & Harmon-Jones, 2008; Payne, 2005), group demographics (e.g., Gonsalkorale, Sherman, & Klauer, 2009), performance on explicit measures (e.g., Klauer et al., 2007; Payne, 2001), performance on different implicit measures (e.g., Payne, 2005), neuropsychological measures (e.g., Amodio et al., 2008; Beer et al., 2008), social judgments (e.g., Payne, 2005), and behavior (e.g., Gonsalkorale, von Hippel, Sherman, & Klauer, 2009). Model-derived process estimates may predict these outcomes with greater specificity and acuity than raw implicit task performance. As an example, model components derived from two different implicit attitude measures (e.g., estimates of association activation derived from two different measures) may correlate more strongly with one another than the behavioral biases (e.g., based on reaction times or errors) demonstrated on the two measures (e.g., Payne, 2005). Likewise, model components derived from an implicit measure of an attitude may correlate more strongly or subtly than the reaction time–based behavioral bias with an explicit measure of the attitude or with an attitude-relevant behavior (e.g., intergroup behavior; Gonsalkorale, Sherman, et al., 2009). In this way, modeling helps to clarify the meanings of the different measures and provides an enhanced method for determining the extent to which different measures (or a measure and a behavior) reflect the same or different underlying component processes and representations. This process also provides a means of enhancing the demonstrated relationships among measures and behaviors that may appear to lack correspondence. These issues are addressed later in greater detail.

The component process estimates provided by models also may be used as dependent variables. For example, one might examine the impact of exemplar exposure (e.g., Gonsalkorale, Allen, Sherman, & Klauer, 2009), social context (Conrey et al., 2005; Lambert et al., 2003); stimulus context (e.g., Allen, Sherman, & Klauer, 2009), processing goals (Payne, Lambert, & Jacoby, 2002), or training programs (e.g., Plant, Peruche, & Butz, 2005; Sherman et al., 2008) on specific process estimates derived from modeling an implicit measure. As described in more detail later, this permits a better understanding of the specific means by which these manipulations influence performance on implicit measures (see Gawronski & Sritharan, Chapter 12, this volume).

It is important to note that applying or fitting a model provides estimates of the process parameters but, in itself, cannot validate the psychological meanings of those parameters. That is, whether the parameters reflect their intended processes must be established independently via construct validation studies. For example, if a parameter is meant to reflect an automatic association or process, then that automaticity must be demonstrated empirically. This might be achieved by showing that the influence of the underlying association or process on the parameter estimate in question is unaffected by a cognitive load or a short response deadline (i.e., the efficiency component of automaticity). In a similar way, the qualitative, psychological nature of a parameter must be established. For example, if a parameter is meant to represent the process of overcoming activated associations, this might be demonstrated by showing that estimates of the parameter are diminished among older adults (for whom self-regulation is diminished) or are associated with reduced expressions of behavioral bias in a one-on-one intergroup interaction (e.g., Sherman et al., 2008).

The equations that describe the relationships among parameters in a model may be consistent with many different psychological meanings of those parameters. For example, the equations in the Control Default model of PD (Payne, 2001) stipulate that automatic race bias influences responses only if controlled detection of the correct response fails (see later discussion). However, given this mathematical constraint, there are a number of potentially valid ways to describe the psychological meanings of the automatic and controlled components (Klauer & Voss, 2008). For instance, the automatic and controlled parameters may reflect processes that operate in parallel, with the controlled process resolving conflicts between simultaneously generated automatic biases and correct responses. Alternatively, the automatic process may generate an initial response that is either corrected or not by a subsequent controlled process. Another possibility is that the automatic process is engaged only after controlled processing fails to provide a response. Each of these possible accounts is consistent with the equations of the Control Default model and the requirement that the automatic component drives behavior only when control fails. The same multiplicity of processing accounts exists for all multinomial models, including the Quad model (Sherman et al., 2008) and the ABC model (Stahl & Degner, 2007). The only way to distinguish among different processing accounts and to determine the psychological meaning of the parameters is through careful validation studies.

We now describe some of the methodological difficulties surrounding the use of implicit measures and the corresponding advantages of modeling approaches.

DIFFICULTIES WITH THE USE AND INTERPRETATION OF IMPLICIT MEASURES

People may be unaware of their underlying knowledge and evaluations or unwilling to report them truthfully. These "willing and able" issues are two of the most difficult problems for research on attitudes and social cognition. Implicit measures were developed, in part, to avoid these obstacles by measuring attitudes and knowledge without directly requesting that respondents report them. In many cases, people are unaware that these constructs 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 frequently seen as reflecting the unintended, automatic activation of stored knowledge, 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 conjunction with explicit measures (e.g., questionnaires), implicit measures are often used to compare and contrast automatic and controlled (or implicit and explicit) social cognition.

Task Confounds

Although this task dissociation approach has proven useful, it has significant drawbacks. First, assessing automatic and controlled components of knowledge with separate measures introduces a confound between process type (e.g., automatic vs. controlled) and measurement task (e.g., Implicit Association Test vs. questionnaire). Undoubtedly, implicit measures are less subject to 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. Indeed, when these task-specific differences are reduced, correlations between implicit and explicit measures are increased (Gawronski & LeBel, 2008; Payne, Burkley, & Stokes, 2008; Ranganath, Smith, & Nosek, 2008).

An instructive example of the possible significance of confounding task and process can be found in the memory literature. As in the domain of implicit social cognition, 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 overarching reinterpretation of dissociations between implicit and explicit measures of social cognition. However, Roediger's example should serve as a cautionary tale for applying task dissociation logic to implicit social cognition.

Implicit Measures Are Not Process Pure

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; he or she simply responds to the color of the ink.

The same principle applies to implicit measures in social cognition, 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 (for a review, see Gawronski, Deutsch, LeBel, & Peters, 2008). Thus, the performance of two people who appear to have different implicit attitudes may reflect nonattitudinal processes. For example, the two people may have equally strong evaluative associations, but one person is better able to effectively overcome them in responding to the demands of the task. In the same way, 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 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., Sherman et al., 2008).

Simultaneous Processes

A related drawback to the task dissociation approach is that it cannot reveal the simultaneous contributions of multiple-component processes, both automatic and controlled. If we assume that responses on any implicit measure reflect the joint contributions of automatic and controlled processes (or even multiple automatic and controlled processes), then it would be advantageous to have a means to track those contributions independently. However, because implicit and explicit measures often are taken as self-contained, process-pure estimates of a single, specific automatic or controlled component, there is no way to assess the ongoing interplay of these processes in producing a discrete response on a particular task. This necessarily produces an overly simplified depiction of the processes that underlie performance on implicit measures.

Implicit Social Cognition Is Constructed, Not Revealed

The preceding discussions all converge on the important point that responses on implicit measures are just that: responses on measures. As such, there are many factors and processes that may intervene in the translation of underlying representations into responses on implicit measures. Although the constructive nature of responses on explicit measures of attitudes has been well appreciated (e.g., Wilson & Hodges, 1992), generally, the same has not been true for implicit measures. Because responses on implicit measures typically are 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).

However, although implicit measures are certainly less susceptible to intention and less reliant on awareness than are explicit measures, 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; Sherman et al., 2008). Moreover, implicit measures of attitudes show poorer test–retest reliability than do explicit measures and 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 that are stable across contexts. Furthermore, 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.

Advantages of Modeling

Modeling approaches avoid these difficulties, assuming that no measure is process pure and that multiple processes (both automatic and controlled) exert independent and simultaneous influences on any task. The purpose of modeling is to identify and quantify these processes, describe how they interact to produce observed outcomes on the measures (e.g., biases in error rates or reaction times), and describe how they relate to other measures and behaviors. Because estimates of the component processes are derived from behavior on a single task, modeling techniques also avoid confounding task and process.

MODELS OF IMPLICIT MEASURES

We now review formal models that have been proposed to account for performance on different implicit measures. This section is not meant to offer detailed descriptions of the procedures involved in applying the models. For that level of detail, interested readers should seek out the primary research articles that proposed and tested these models, and which are cited in this chapter.

Signal Detection (Green & Swets, 1966), PD, and multinomial models (Batchelder & Riefer, 1999) have been used to model the accuracy data from implicit measures of social cognition. The multinomial models include variations of PD models (Bishara & Payne, 2009) and extensions of those models, including the Quad model (Sherman et al., 2008) and the ABC model (Stahl & Degner, 2007). Finally, Klauer and his colleagues (2007) have developed a diffusion model that takes into account both error rates and reaction times in accounting for performance on the IAT.

Signal Detection Theory

SDT has been used primarily to account for two-option decisions in the domains of perception and memory (e.g., Green & Swets, 1966). More recently, it also has been applied to understanding responses on the Shooter task, an implicit measure of stereotyping (e.g., Correll et al., 2002; Greenwald, Oakes, & Hoffman, 2003). To our knowledge, it has not been applied to other implicit measures, although it is equipped to analyze data from the Weapons task (e.g., Payne, 2001) and other semantic and evaluative priming tasks.

The purpose of SDT is to separate sensitivity (or discrimination accuracy) and response bias in responding. For example, in the domain of recognition memory, SDT identifies one response component that reflects the extent to which perceivers are able to discriminate between items they have and have not encountered and a second component that reflects a bias to call an item old or new. Although many different equations have been offered to measure sensitivity (for a review, see Snodgrass & Corwin, 1988), all of them propose that it is reflected by a difference in the prevalence of hits (i.e., correctly calling an old item *old*) versus false alarms (i.e., incorrectly calling a new item *old*). By contrast, response bias has been defined by a number of alternative relationships between hits and false alarms (for reviews, see Macmillan & Creelman, 1990; Snodgrass & Corwin, 1988). An important point about SDT in the context of implicit social cognition is that the component processes are not assumed to map onto the distinction between automatic and controlled processes.

In the Shooter task, SDT has been used to separate people's ability to discriminate accurately whether a target is holding a gun or another object from a bias to provide a gun response (shoot) versus no-gun response (don't shoot). Correll and colleagues (2002; see also Correll et al., 2007) reported that participants set a lower response criterion for reporting the presence of a gun for black targets than for white targets. However, the ability to discriminate accurately the presence of a gun did not differ between black and white targets. In contrast, Greenwald and colleagues (2003) found both the criterion effect reported by Correll and colleagues and also that participants were better able to discriminate between guns and non-guns for white targets than black targets. There are a variety of procedural differences in the studies by Correll and colleagues (2002, 2007) and Greenwald and colleagues, and the basis for the discrepant results is not clear.

Process Dissociation

Concerns about task confounds and assumptions of process purity in the literature on implicit versus explicit memory led Jacoby (1991; Lindsay & Jacoby, 1994; see Jacoby, Kelley, & McElree, 1999, for a review) to develop PD techniques for separating the automatic and controlled components of memory 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.

Payne (2001) recognized that the same issues that are problematic for separating implicit and explicit memory also are problematic for separating implicit and explicit social cognition, and adapted PD techniques for use in decomposing automatic and controlled components of implicit task performance. This technique has been applied primarily to the Weapons task (for a review, see Payne, 2008) but also to the Shooter task (e.g., Plant et al., 2005) and the IAT (e.g., Conrey et al., 2005; Huntsinger, Sinclair, & Clore, 2009; Sherman et al., 2008; Stewart, von Hippel, & Radvansky, 2009). In principle, PD is applicable to any task that compares compatible (e.g., black + bad/white + good) and incompatible (e.g., black + good/white + bad) trials, in which automatic and controlled processes are placed in concert with and in opposition to one another, including many varieties of semantic and evaluative priming tasks, the IAT, the Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001), and the Extrinsic Affective Simon Task (EAST; De Houwer, 2003).

Initially, Jacoby and his colleagues developed two different models of PD (Jacoby, 1991; Lindsay & Jacoby, 1994). The primary difference between the two models is whether automatic or controlled processes are assumed to be dominant.

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 (1991) 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. This is the PD model that Payne and others have applied to the Weapons task and the Shooter task (e.g., for a review, see Payne, 2008; Plant et al., 2005).

STRUCTURE AND USE OF THE CONTROL DEFAULT MODEL

The logic of the model for these and other priming tasks is as follows (see Figure 9.1). On compatible trials (e.g., a black prime followed by a weapon target), participants may produce the correct response either by correctly identifying the target through controlled processing (C) or by relying

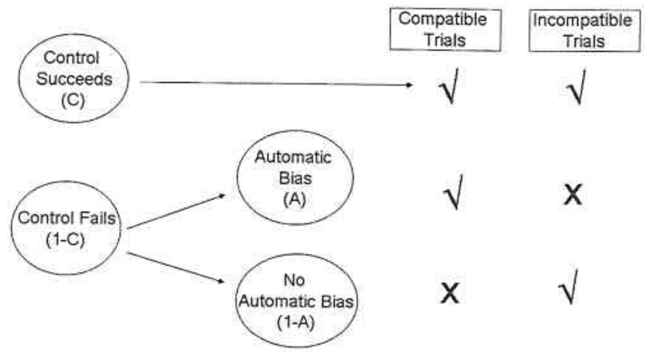

FIGURE 9.1. The Control Default process dissociation model. Each path represents a likelihood. Parameters with lines leading to them are conditional upon all preceding parameters. The table on the right side depicts correct (✓) and incorrect (×) responses as a function of process pattern and trial type.

on an automatic stereotypic bias (A) to produce a "weapon" response, when controlled detection of the object fails ($A^*(1 - C)$). Thus, the equation for correct compatible responses is $C + A\ (1 - C)$. In contrast, on incompatible trials (e.g., a black prime followed by a tool target), participants will produce an incorrect response only when controlled detection fails, and participants rely on an automatic bias ($A^*(1 - C)$). Subtracting the equation for incompatible trials, in which automatic and controlled processes produce opposite responses, from the equation for compatible trials, in which automatic and controlled processes produce the same response, provides an estimate of controlled processing, C. Subsequently, an estimate of automatic processing (A) can be derived through simple algebra.[1]

As indicated by the equations, this version of PD proposes that the race of the prime influences judgments about the target object only if people fail to identify the target object through controlled discrimination processes. If the object can be correctly identified, then race will have no influence on judgments. Thus, the model permits no role for automatic processes (e.g., activation of stereotypic associations) that capture attention and influence behavior even though the correct response can be determined. Researchers applying this model must be careful to interpret the automatic component of this model accordingly.

VALIDATION OF THE CONTROL DEFAULT MODEL

A multinomial model implementation has shown that the Control Default model accurately predicts performance on the Weapons task, to which it has been applied primarily (Bishara & Payne, 2009; Sherman et al., 2008). The parameters also have been shown to vary independently of one another. For example, implementing a response deadline reduced C but left A unaffected (e.g., Payne, 2001). This finding also provides evidence of the controlled and automatic natures of C and A, respectively. Other findings further support this conclusion. For example, depleting processing resources through a lengthy ego-depletion manipulation reduced C but not A (Govorun & Payne, 2006). In yet another study, A correlated with responses on implicit measures of bias, whereas C correlated with other measures of cognitive control such as an antisaccade task (Payne, 2005). Finally, Stewart and colleagues (2009) showed that C but not A was associated with age-related deficits in inhibitory ability.

The specific qualitative natures of the A and C parameters are somewhat less clear. The status of the A parameter as representing a race-based bias is supported by its correlations with behavioral bias on implicit measures. However, the underlying basis for this A bias is unspecified. The C parameter appears to be associated with a variety of controlled processes. One study showed that the C parameter correlated with performance on an antisaccade task, which assesses the ability to inhibit attention from being directed to a distracting cue (Payne, 2005). The just-mentioned data reported by Stewart and colleagues (2009) also shows that C is associated with inhibition ability. Other studies have shown that C correlates with motivations to control biased responding (Amodio et al., 2008; Payne, 2005). Still other data show that the C parameter correlates with event-related brain potentials that have been linked to monitoring conflicting response tendencies (Amodio et al., 2004, 2008). The detection of such conflict assists in the subsequent control of unwanted responses. Although these different indices of control are related, they also point to distinct processes. The C parameter may reflect each of these processes and more (e.g., Klauer & Voss, 2008). One challenge for future research will be to specify the qualitative natures of the PD component processes.

APPLICATION OF THE CONTROL DEFAULT MODEL

Predicting Other Measures and Behavior. One application of the Control Default model has been to use the A and C parameters derived from Weapons task performance to predict responses on other measures of racial bias. In one study, the A parameter was shown to correlate with an explicit measure of bias only among those participants low in motivation to control prejudiced responses (Payne, 2001). This demonstrates the role of motivation in obscuring bias on explicit measures. Another study showed that the A parameter better predicted performance on both a Weapons task and an evaluative priming task (e.g., Fazio et al., 1995) among participants with lower levels of the C parameter (Payne, 2005). Importantly, the C estimates were derived from different tasks than the Weapons task and priming task. Thus, control, as measured on a separate task, predicted the extent to which automatic bias influenced implicit task performance. This demonstrates the role of control in implicit task performance. The same relationship between A and C was found in predicting stereotyping in an impression formation task. Payne (2005) also showed that A parameters derived from an evaluative priming task and a Weapons task correlated more strongly than did the behavioral biases (based on errors) demonstrated on the tasks. This shows that weak correlations among different implicit measures may conceal stronger correlations among the component processes of the different measures, and that the weak correlations may be due to aspects of the measures that are not of central interest (i.e., aspects that do not reflect automatic bias).

Accounting for Changes in Implicit Bias. This model also has been used to account for the effects of experimental manipulations on implicit bias. For instance, Lambert and colleagues (2003) found that, surprisingly, racial bias on the Weapons task increased when participants were told that their responses would be observed by others. Application of the Control Default model demonstrated that the public context did not affect the A parameter but reduced C. This suggests that the public context increases bias by interfering with controlled processes that would otherwise prevent biased responses. In another study, Plant and colleagues (2005) showed that training reduced racial bias on the Shooter task. Application of the Control Default model showed that training increased C for both black and white trials but reduced A only for black trials. This suggests that training reduces biased responses by reducing the extent of automatic race bias.

Accounting for Group Differences in Implicit Bias. Finally, Stewart and colleagues (2009) used this model to account for group differences in implicit bias. Specifically, age-related increases in implicit bias were associated with diminished C but were not related to A. At the same time, greater implicit bias among white than black participants was associated with greater A but was unrelated to C. Thus, it appears that aging is associated with greater IAT bias as a result of diminished control, whereas outgroup (vs. ingroup) status is associated with greater IAT bias because of enhanced automatic race bias.

The Automatic Default Model

The other PD model (Lindsay & Jacoby, 1994) was designed to account for tasks in which automatic processes are thought to influence behavior, regardless of whether or not control succeeds. In this model, controlled processes drive responses only in the absence of automatic bias. For example, 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.

The logic of the model for the Weapons task and other priming tasks is as follows (see Figure 9.2). On compatible trials (e.g., a black prime followed by a weapon target), participants may produce the correct response either by relying on an automatic bias to produce a "weapon" response (A) or by correctly identifying the target through controlled processing (C) when no automatic bias is activated $(1 - A)$. Thus, the equation for correct compatible responses is $A + C(1 - A)$. In contrast, on incompatible trials (e.g., a black prime followed by a tool target), this model proposes that a correct response will be provided only if there is no automatic bias and controlled detection succeeds

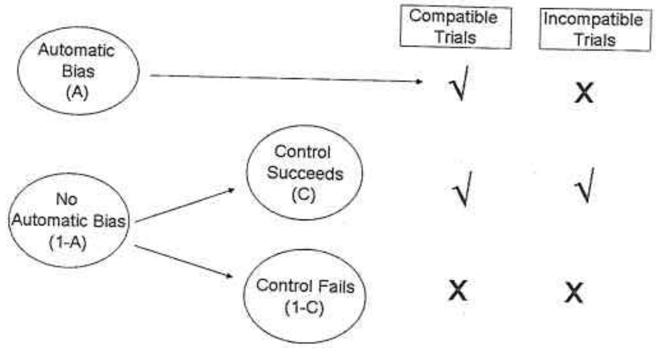

FIGURE 9.2. The Automatic Default process dissociation model. Each path represents a likelihood. Parameters with lines leading to them are conditional upon all preceding parameters. The table on the right side depicts correct (✓) and incorrect (×) responses as a function of process pattern and trial type.

$C^*(1 - A)$. Subtracting the equation for incompatible trials, in which automatic and controlled processes produce opposite responses, from the equation for compatible trials, in which automatic and controlled processes produce the same response, provides an estimate of automatic processing, A. Subsequently, an estimate of controlled processing (C) can be derived through simple algebra.

As indicated by the equations, this version of PD proposes that controlled detection of the target object influences judgments of the target only if the race prime fails to activate an automatic bias. If there is an automatic bias, then control will have no influence on judgments. Thus, the model permits no role for controlled processes that determine responses despite the presence of an automatic bias. As such, this model cannot account for perceivers' ability to produce a correct response on an implicit measure despite the operation of an automatic bias. Researchers applying this model must be careful to interpret the automatic and controlled components of this model accordingly.

One limitation of this model is that it does not distinguish between cases in which an automatic bias is not activated at all from cases in which the bias 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 previously, 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 social cognition, 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 more thorough discussions, see Conrey et al., 2005; Sherman et al., 2008).

Although this model is applicable to all the same tasks as the Control Default model, it has been used very rarely. To our knowledge, it has been used only to model the Weapons task, and it has failed to provide adequate fit for task performance (e.g., Bishara & Payne, 2009), perhaps because of the limited role the model affords to control. Nevertheless, the limited use of this model is somewhat surprising, given that the automatic parameter in this model appears to be more akin to what researchers usually mean when they refer to automatic attitudes or stereotypes. For example, in most dual-process models of social cognition (Chaiken & Trope, 1999), the role of the automatic process is to influence perception and behavior regardless of the status of controlled processes. Controlled processes may moderate the impact of the automatic process, but they do not preclude its influence (for a review, see Sherman et al., 2008). Yet the PD model used almost exclusively is the Control Default model, which assigns a secondary role to automatic processes, such that their influence is felt only when there is no controlled processing.

The Quad Model

The Quad model was developed as an extension of PD (Conrey et al., 2005; Sherman et al., 2008). Whereas the PD models estimate a single automatic and a single controlled process, the Quad model estimates four processes, two of which are meant to reflect features of automatic processing and two of which are meant to reflect features of more controlled processing. In addition to providing estimates of these four qualitatively distinct processes, the other primary purpose of the Quad model is to separate cases in which an automatic association is activated but overcome from cases in which the association simply is not activated. This is the problem described earlier in reference to adult versus child performance on the Stroop task, and it is not addressed by either PD model.

Different dual-process models of social cognition have proposed qualitatively distinct automatic and controlled processes (for a review, see Sherman et al., 2008). Most commonly, automaticity is represented as simple associations that are triggered by the environment without the perceiver's awareness or intent and that influence subsequent processing (e.g., Fazio et al., 1995; Greenwald et al., 1998). This is the kind of automaticity to which researchers studying implicit social cognition typically refer, and it is assessed in the Automatic Default PD model. In other dual-process models (e.g., Jacoby, 1991), however, automatic processes influence behavior only when control fails. This is the role of automaticity in the Control Default PD model.

Dual-process models also have generally been concerned with one of two different types of control. In some models (e.g., models of impression formation or persuasion), control is characterized by stimulus-detection processes that attempt to provide an accurate depiction of the environment (e.g., Brewer & Feinstein, 1999; Chen & Chaiken, 1999; Fiske, Lin, & Neuberg, 1999; Petty & Wegener, 1999). This is how the C component has been described in both 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 (see also Wegner, 1994).

The Quad model is a multinomial model (see Batchelder & Riefer, 1999) designed to measure the contributions of each of these processes to performance on implicit measures of social cognition. More formally, the four components of the model are (1) the activation of an association (Association aCtivation [AC]), (2) the ability to determine correct and incorrect responses (Detection [D]), (3) the success at overcoming activated associations when necessary (Overcoming Bias [OB]), and (4) 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 meant to reflect features of automatic processing (although G need not; see later discussion), D and OB are meant to reflect features of controlled processing. The relatively automatic versus controlled nature of these processing components has been supported by empirical research (see later discussion).

Structure and Use of the Quad Model

The structure of the Quad model is depicted as a processing tree in Figure 9.3. 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

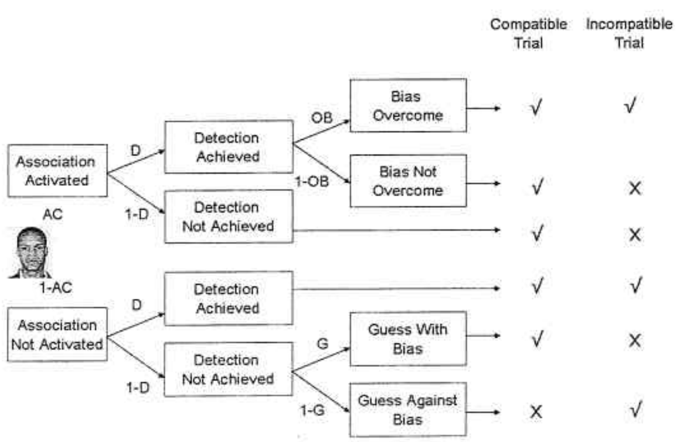

FIGURE 9.3. The Quadruple Process model. Each path represents a likelihood. Parameters with lines leading to them are conditional upon all preceding parameters. The table on the right side depicts correct (✓) and incorrect (×) responses as a function of process pattern and trial type.

a black face may activate negative associations (AC), 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 bias will be either congruent or incongruent with the correct answer "black" achieved through detection (D). On compatible trials, there is no conflict between activated associations and what is detected. As such, there is no need to overcome the bias (OB) 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 (G). Guessing need not be random and may be quite strategic (rather than automatic), such as a bias to respond with the positive key when all else fails (Conrey et al., 2005). Other response biases reflect more automatic processes, such as a dominant-hand bias or a bias to press the button with the higher probability of providing a correct response (Conrey et al., 2005).

Parameter estimates cannot be generated with simple algebra but must be established via maximum likelihood estimation. In Figure 9.3, each path represents a likelihood. Processing parameters with lines leading to them are conditional upon all preceding parameters. For instance, OB is conditional upon both AC and D. Similarly, G is conditional upon the lack of AC (1 – AC) and the lack of D (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, the activation of associations (AC), attempts to detect a correct response (D), and attempts to overcome associations (OB) may occur simultaneously. However, in determining a response on an incompatible trial, the status of OB determines whether AC or D drives responses.

The conditional relationships described by the model form a system of equations that predict the number of correct and incorrect responses in different conditions (e.g., compatible and incompatible trials). For example, a black face on an incompatible trial will be responded to correctly with the probability: $AC \times D \times OB + (1 - AC) \times D + (1 - AC) \times (1 - D) \times (1 - G)$. This equation sums the three possible paths by which a correct answer can be returned in this case. The first part of the equation, $AC \times D \times OB$, is the likelihood that the association between black and negative is activated, that the correct answer can be detected, *and* that the association is overcome in favor of the detected response. The second part of the equation, $(1 - AC) \times D$, is the likelihood that the association is not activated *and* that the correct response can be detected. Finally, $(1 - AC) \times (1 - D) \times (1 - G)$ is the likelihood that the association is not activated, that the correct answer cannot be detected, *and* that the participant guesses correctly. The respective equations for each item category (e.g., black faces, white faces, positive words, and negative words in both compatible and incompatible blocks) are then used to predict the observed proportions of errors in a given data set. The model's predictions are then compared with the actual data to determine the model's ability to account for the data. A chi-square estimate is computed for the difference between the predicted and observed errors. To best approximate the model to the data, the four parameter values are changed through maximum likelihood estimation until they produce a minimum possible value of the chi-square. The final parameter values that result from this process are interpreted as relative levels of the four processes.

Although we have used the IAT as an example, 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, including Stroop tasks, evaluative priming tasks (e.g., Allen et al., 2009; Conrey et al., 2005; Sherman et al., 2008), the Weapons task (e.g., Conrey et al., 2005), the Shooter task, and the GNAT (Gonsalkorale, von Hippel, et al., 2009).

Comparisons to PD Models

The Quad model is an extension of PD models and includes features of both the Control Default and Automatic Default models. In particular, the AC and D parameters in the Quad model are analogous to the A and C parameters of the PD models in terms of their qualitative features (although these features are specified in greater detail in the Quad model). However, the hypothesized relationships among the components differ in important ways in PD and Quad models. PD models posit that automatic and controlled processes either dominate or are dominated by one another. In the Control Default model, control (C) dominates; in the Automatic Default model, automaticity (A) dominates. In contrast, the Quad model suggests that either conflicting AC or D processes may act as the default, depending on the outcome of the OB process (another controlled process). When there is a conflict between AC and D (i.e., on incompatible trials), AC dominates D when OB fails, and D dominates AC when OB succeeds. Thus, whereas the final estimates of A and C in the PD models reflect only situations in which one or the other process has dominated, estimates of AC and D in the Quad model reflect cases in which the two parameters both were superordinate and subordinate to one another. That is, dominating or being dominated are defining features of automaticity and control in PD models, whereas the Quad model does not limit the influence of automatic and controlled processes in this way.

Of course, the Quad model also differs from PD models in the specification of the OB and G parameters. OB represents a form of controlled process that is common in the dual-process literature but is not represented as a distinct process in either of the PD models. It plays a critical role in the Quad model, acting as the arbiter of whether AC or D determines responses when they are in conflict. When successful, OB inhibits incorrect association-based responses in favor of detected correct responses.

Similarly, the G parameter in the Quad model represents a form of response bias that is not represented as a distinct process in PD models. When there is no automatic activation of a biased response or controlled detection of a correct response, G determines the response that will be given. The addition of the G parameter in the Quad model implies very different outcomes than the PD models when both the automatic activation process and the controlled detection process fail. In the Control Default model, this circumstance leads to correct responses on compatible trials and incorrect responses on incompatible trials. That is, in the absence of all information, the default is to produce a response that is correct on compatible trials and incorrect on incompatible trials. In the Automatic Default model, this same circumstance always leads to incorrect responses. That is, if there is no automatic activation and no controlled detection, an incorrect response will always be provided on all trials. In contrast, the G parameter of the Quad model proposes that, in the absence of activation and detection processes, a guessing process occurs that will sometimes produce correct and sometimes produce incorrect responses.

Validation of the Quad Model

The Quad model has shown its ability to predict performance accurately on a variety of priming tasks, IATs, and the GNAT, demonstrating good model fit for these tasks (Conrey et al., 2005; Gonsalkorale, von Hippel, et al., 2009; 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). Aging was associated with an increased ability to detect stimuli accurately (D) but decreased success at overcoming bias (OB; Gonsalkorale, Sherman, et al., 2009). This age-based dissociation between OB and D is consistent with previous research (Rosano et al., 2005) showing that not all forms of controlled processing diminish with age. Rather, age-related deficits in cognitive control appear to be related primarily to self-regulatory processes. 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 also has been established by a number of findings (Conrey et al., 2005; Sherman et al., 2008). For example, 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.

Beyond the general automatic/controlled distinction, the specific qualitative nature of the different parameters also has been established. The status of AC as a measure of association activation is supported by the fact that AC was shown to be positively correlated with reaction time bias on the IAT (Conrey et al., 2005) and with activation of the amygdala and insula in a neuroimaging study (Beer et al., 2008). The amygdala and insula are known to be involved in emotional processing and arousal. The validity of OB as a measure of self-regulation is supported by demonstrations that it is impaired by alcohol consumption and decreases with age, two factors associated with impairments in self-regulation (Gonsalkorale, Sherman, et al., 2009; Sherman et al., 2008). OB also has been shown to be negatively correlated with reaction time bias on the IAT (Conrey et al., 2005) and positively correlated with favorable intergroup interactions (Gonsalkorale, von Hippel, et al., 2009), further evidence that OB measures the inhibition of bias. 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. The precise qualitative nature of the D parameter is less well established. Neuroimaging data showed that 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. However, further research is needed to establish the precise psychological meaning of D.

Application of the Quad Model

MALLEABILITY OF RESPONSES ON IMPLICIT MEASURES

The Quad model has been applied to a number of empirical and theoretical problems surrounding implicit social cognition. One set of questions pertains to understanding the contextual malleability of responses on implicit measures. Although they originally were assumed to be highly stable and resistant to change, considerable research now indicates that responses on these measures are highly context dependent (for reviews, see Blair, 2002; Gawronski & Bodenhausen, 2006; Sherman et al., 2008). When responses change as a result of situational manipulations, what accounts for it? Most typically, these effects are attributed to context-dependent changes in the specific associations that are activated in the different contexts (e.g., Blair, 2002; Ferguson & Bargh, 2007; Gawronski & Bodenhausen, 2006) or to changes in the degree to which given associations are activated in the different contexts (e.g., Glaser & Knowles, 2007; Maddux, Barden, Brewer, & Petty, 2005; Moskowitz, Gollwitzer, Wasel, & Schaal, 1999; see also Klauer, 2009, for a different view). According to the Quad model, however, such effects could be due to changes in the nature of the activated associations, changes in respondents' ability to determine appropriate behavior, changes in respondents' ability to overcome associations when necessary, changes in response biases, or some combination of these processes.

In some cases, application of the Quad model suggests that changes in implicit task performance primarily reflect changes in the underlying associations that are activated (either the extent of activation or the nature of the associations activated). For example, when newly formed implicit attitudes about a target person are gradually altered (e.g., Rydell & McConnell, 2006), the effect is associated only with changes to the AC parameter (Sherman et al., 2008). Similarly, the well-replicated finding that the presentation of prejudice-inconsistent exemplars reduces implicit bias (e.g., Blair, Ma, & Lenton, 2001; Dasgupta & Greenwald, 2001; Govan & Williams, 2004; Mitchell, Nosek, & Banaji, 2003) is related solely to changes in the associations that are activated (AC; Gonsalkorale, Allen, et al., 2009; Sherman et al., 2008).

Other interventions that change implicit bias appear to influence both automatic and controlled processes. For example, training to negate stereotypes has been shown to reduce subsequent implicit stereotyping (e.g., Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000). A Quad model analysis of this effect showed that the negation training both reduced AC and increased the ability to detect correct responses (D). So training not only altered the associations that were activated, but improved participants' ability to perform the task accurately (Sherman et al., 2008). A reanalysis of Lambert and colleagues' data (2003) on the effects of public accountability on implicit stereotyping showed that, relative to the private condition, the public condition reduced D (this is similar to the C effect reported by Lambert et al.) and increased AC. Thus, unlike the PD analysis of these data, the Quad model analysis suggested that public accountability not only diminished control but also increased the activation of biased associations, supporting a social facilitation account of the finding, in which the dominant response (in this case, stereotype activation) is enhanced by the presence of others (Zajonc, 1965; see also Conrey et al., 2005).

In still other cases, variations in implicit bias appear to have nothing to do with the underlying associations but rather reflect only variations in controlled processes. Bartholow, Dickter, and Sestir (2006) found that participants under the influence of alcohol demonstrated greater implicit stereotyping than their sober counterparts. A Quad model analysis of these data showed that alcohol reduced estimates of overcoming bias (OB) but had no effect on the activation of associations (Sherman et al., 2008). This finding is consistent with research showing that alcohol impairs self-regulatory ability. Another example involves the effects of stimulus contexts on implicit prejudice. A number of studies have shown that implicit prejudice is reduced when pictures of black and white targets are presented in prejudice-inconsistent contexts (e.g., black targets in front of a church; white targets in front of a jail). Allen and his colleagues replicated this finding in both an IAT and an evaluative priming task (Allen et al., 2009). In both cases, application of the Quad model showed that the inconsistent contexts increased OB but affected no other parameters. Thus, it appears that these contexts act as cues to initiate self-regulatory processes that overcome biased associations.

VARIABILITY IN RESPONSES ON IMPLICIT MEASURES

Another set of questions surrounds the meaning of individual and group differences in implicit task performance. To what extent do these differences reflect variation in underlying associations, the ability to determine appropriate behavior, the ability to overcome associations, response biases, or some combination of these processes? As is the case for malleability effects, variability effects sometimes are related to the automatic activation of associations, sometimes are related to a combination of automatic and controlled processes, and sometimes are related only to control.

Not surprisingly, black people show less positivity toward whites and negativity toward blacks on implicit measures than do white people (e.g., Nosek, Banaji, & Greenwald, 2002). In one study, application of the Quad model showed that only the AC parameter differed between black and white participants (Gonsalkorale, Allen, et al., 2009). Thus, black people are not better at controlling their racial bias, they just have less biased associations in the first place, perhaps because of more frequent exposure to positive black exemplars.

Internal and external motivations to respond in nonprejudiced ways have been shown to be important moderators of the extent of implicit bias (Amodio et al., 2008; Amodio, Harmon-Jones, & Devine, 2003; Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002). Specifically, individuals who are internally but not externally motivated (high Internal Motivation to Respond Without Prejudice Scale [IMS]/low External Motivation to Respond Without Prejudice Scale [EMS]) to behave in nonprejudiced ways demonstrate less bias on measures of implicit bias than individuals who are motivated by both internal and external reasons (high IMS/high EMS) or who lack internal motivation (low IMS). Quad model analyses of both Weapons task data collected by Amodio and colleagues (2008) and a new IAT study showed that high-IMS/low-EMS participants exhibited less activation of biased associations (AC) and a greater likelihood of detecting correct responses (D) than other participants (Sherman et al., 2008). Thus, like participants who are directly trained at overcoming bias, these individuals (who are believed to train themselves to act in nonbiased ways; Monteith, Ashburn-Nardo, Voils, & Czopp, 2002) have weaker associations and stronger detection of appropriate responses.

Other individual differences in implicit bias appear to have nothing to do with activated associations and are, seemingly, based entirely on variations in controlled processes. Recent research has revealed a developmental trend, showing a positive correlation between age and implicit racial bias among white people (e.g., Nosek et al., 2002). This finding often is interpreted as evidence that older people's racial associations are more biased than those of younger adults, reflecting generational changes in societal attitudes. However, an alternative explanation for age differences in prejudice is that deficits in self-regulatory ability alter the attitudinal expression of older adults on implicit measures. Indeed, application of the Quad model to IAT data showed that the increase in IAT bias with age was associated only with a decreased ability to overcome bias (OB; Gonsalkorale, Sherman, et al., 2009). Thus, age differences in implicit racial bias appear to be due to age-related losses in regulatory functions.

PREDICTING BEHAVIOR

Finally, the Quad model also has been applied toward understanding the underlying automatic and controlled processes that predict social behavior. A recently published study used the Quad model to examine the processes that predict the quality of a social interaction between members of different social groups (Gonsalkorale, von Hippel, et al., 2009). White non-Muslims interacted with a Muslim confederate and completed a GNAT measuring anti-Muslim bias. The confederate's ratings of how much he liked the interaction partners were predicted by an interaction between AC and OB. 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 negative associations on the GNAT predicted the quality of the social interaction when those associations were strong.

The ABC Model

The ABC model was developed by Stahl and Degner (2007) to account for performance on the EAST task specifically (De Houwer, 2003). The ABC model is an extension of the Automatic Default PD model, in that it assumes that the automatic component (A) will drive responses if it is engaged. The controlled component (C) determines responses only if A is not engaged. The ABC model extends the Automatic Default model by adding a parameter that measures guessing that occurs when neither A nor C are engaged. Thus, the ABC model proposes that, in the absence of all other relevant information, a guess will be made that will yield either a correct or an incorrect response. This is in contrast to the assumption of the Automatic Default model that an incorrect response will always result when neither A nor C are engaged.

In four studies, Stahl and Degner (2007) showed that the ABC model provided excellent fit for EAST data. They also showed that the A parameter corresponded to the automatic activation of evaluative associations, whereas the C parameter reflected the use of controlled processes to respond to task demands. As predicted, the A parameter was sensitive to variations in the evaluative nature of the stimuli that were irrelevant to task performance but not to variations in the difficulty of responding to the task-relevant feature (reflecting the automatic nature of the evaluative process). In contrast, the C parameter was sensitive to the difficulty of task-relevant demands but not to the task-irrelevant evaluative features of the stimulus. Finally, they showed that only the guessing parameter (B) was influenced by a manipulation of the baseline proportions of responses requiring different keys. Specifically, participants were more likely to guess with the key most likely to produce a correct response.

The Diffusion Model of the IAT

Recently, Klauer and his colleagues developed a diffusion model to account for IAT performance (Klauer et al., 2007; see also Brendl, Markman, & Messner, 2001, for a related random-walk model that was proposed but not formally tested). Diffusion models differ from SDT, PD, and multinomial models in that the models' parameters are estimated from both error rates and reaction times. Given that implicit bias is most typically presented in terms of reaction times, this is a distinct advantage of the Diffusion model.

The Diffusion model assumes that the choice between the two responses on an IAT trial is based on an accumulation of information over time (see Figure 9.4). The diffusion process moves from a starting point between the two possible responses until one of the response thresholds is reached and the response associated with it is initiated. The model estimates seven parameters that contribute to the ultimate response. Parameter a describes the amount of information that must accumulate before a decision is made. Therefore, it represents speed–accuracy trade-off settings. Parameter z is the starting point of information accumulation and measures response bias. For example, a starting point close to the upper threshold implies that comparatively little additional information must accumulate toward the upper threshold before it is crossed; conversely, comparatively more information must accumulate toward the lower threshold before it can be crossed. The result is a response bias toward the response associated with the upper threshold and against the response associated with the lower threshold. Parameter v is the mean drift rate. Drift rate quantifies the direction (toward lower vs. upper threshold) and speed with which relevant information accumulates. A high drift rate implies both fast and accurate decisions. Parameter t_0 represents the contribution of nondecision processes relating to, for example, preparatory encoding of stimuli and motor responses. Finally, three parameters measure variability in the drift rate (η), variability in the starting point (s_z), and variability in the nondecision processes (s_t). The parameters can be estimated via chi-square and weighted least squares methods that rely on grouped data or on the maximum-likelihood method that relies on ungrouped latencies. Overall model fit is assessed with a chi-square distributed goodness-of-fit statistic (for a complete description, see Klauer et al., 2007).

In an initial application of the model, Klauer and colleagues (2007) showed that the model provided adequate fit to IAT data. They also found that the speed–accuracy trade-off was more conservative in incompatible (e.g., flowers–bad; insects–good) than compatible trials, a prediction that had been offered by Brendl and colleagues (2001) but not directly tested. Results also showed a slower drift rate (i.e., accumulation of information) and slower nondecision components in incompatible than compatible trials. However, only the differential drift rate in compatible and incompatible trials predicted responses on an explicit measure of attitudes, suggesting that this parameter may be particularly indicative of attitudinal responses. In addition, results showed that the speed–accuracy trade-off difference on compatible and incompatible trials reflected method variance rather than attitude-specific responses. This suggests that unwanted sources of variance in the IAT related to factors that influence speed–accuracy trade-offs, such as age, prevention versus promotion focus, instructions, strategies, processing styles, and so on may influence results.

Choosing among Models

Although each of these models has been applied primarily (or only) to one specific implicit measure, each also may be applied to other measures (assuming that the model provides adequate fit for the data). How then should one choose which model to apply? In large part, the answer to this question depends on the purpose for which one is using the model. If the purpose of modeling is to find the set of processes (and relationships among them) that best describe task performance, then model fit would be the key criterion. However, if the purpose of the modeling is to extract estimates of processes of interest that may then function as distinct variables (e.g., for predicting other variables, measures, or behavior, or for use as dependent variables), then the criteria are more theoretical in nature (Sherman et al., 2008).

Model Fit

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

Theoretical Considerations

If the purpose of modeling is to derive estimates of processes of interest, then the choice of a model should be based on theoretical considerations (assuming that the model provides adequate fit to the data and the meanings of the parameters have been adequately established via construct validation studies). In the same way one may choose to measure attention capture, attentional disengagement, perceptual encoding, conceptual encoding, or any number of other processes in standard behavioral research, when choosing a model, one must decide which processes are most relevant to the research questions at hand. The parameters estimated by each model are, in fact, separate variables representing distinct cognitive processes. Even when they are described in similar ways (e.g., the A parameter in the Control Default model and the AC parameter in the Quad model), there may be critical differences among the processes estimated by different models. Thus, if a researcher is interested in an automatic process that captures attention and influences behavior regardless of whether or not control succeeds, then the Control Default model would not be appropriate. In this model, the A parameter reflects a subordinate automatic process that influences behavior only when control has failed, and the model is not mathematically equipped to estimate a dominant automatic processes. Similarly, if a researcher is interested in speed–accuracy trade-offs, only the Diffusion model will suffice. Hence, in selecting a model, a paramount concern should be which processes are theoretically relevant to the research. If the processes of interest are not represented in existing models, one may always develop and validate a new model.

Why Not Rely Solely on Fit to Choose a Model?

Some may feel that model fit should be the only criterion for choosing a model. Why would a researcher ever use a model other than the one that provides the best fit? First, a strictly atheoretical, bottom-up approach may require one to adopt a nonsensical model. Indeed, from this perspective, a researcher would be bound to test all potential models on each and every data set (even models that may not make theoretical sense). Logically, there is no basis to restrict model-fitting efforts to even a few competing models. This is a recognized problem in all kinds of mathematical modeling of psychological data (e.g., multinomial modeling, SEM, etc.), and is why theoretical plausibility is ascribed such a central role in model building.

Second, in our experience, it is not the case that one model will always provide the absolute best fit to a given measure. For example, we have experimented with five slight variations of the Quad model (e.g., models that have one OB vs. two OB parameters). We have found that, although some variants of the model generally provide better fit than others, this is not the case 100% of the time. Were we to adopt a strict bottom-up approach, we would apply different versions of the model to different data sets based on (often slight) differences in model fit as the only justification. Our experience suggests that the search for a single model or single version of a model that will provide superior fit on all occasions would be a lengthy and fruitless endeavor.

Third, although it is easy to test whether or not a model provides good fit to a data set, the standards for comparing levels of good fit are unclear. How great must the difference in fit be to command use of one model over another? For example, we (Sherman et al., 2008) compared the ability of the Control Default PD model and the Quad model to account for data from priming measures versus the IAT. We found that, overall, both models provided good fit to both measures. However, whereas the Control Default model provided better fit for the priming measures, the Quad model provided better fit for the IAT. Nevertheless, the effect sizes of these differences in model fit were tiny (all Cohen *ds* < 0.017). We conclude that either model may be applied to either task, provided that the model offers adequate fit for the given data set. More generally, we argue that, given that the meaning of a model's parameters have been validated via empirical research and that the model provides adequate fit to the current data set, the primary consideration for choosing the model (vs. another) should be theoretical.

CONCLUSION

Implicit measures of attitudes and knowledge do not provide process-pure estimates of automatic biases. Rather, responses on implicit measures reflect the influence of a variety of automatic and controlled processes. Mathematical modeling of implicit task performance can help to disentangle these component processes and provide independent estimates of their prevalence, without relying on task dissociation techniques that confound processes with measures. In so doing, modeling can further our understanding of implicit social cognition in a number of ways. It can help to clarify the meanings of different implicit measures and distinguish among accounts of how people respond to them. It can clarify the meaning of relationships (or lack of relationships) among different implicit measures, explicit measures, neuropsychological measures, and measures of judgments and behavior and can enhance the strength of these relationships. Modeling also can help to explain variability in implicit measures and the effects of different manipulations on implicit task performance (i.e., task malleability). Finally, it can help researchers conceptualize the nature of automatic and controlled processes, how they are related, and how they interact to produce behavior.

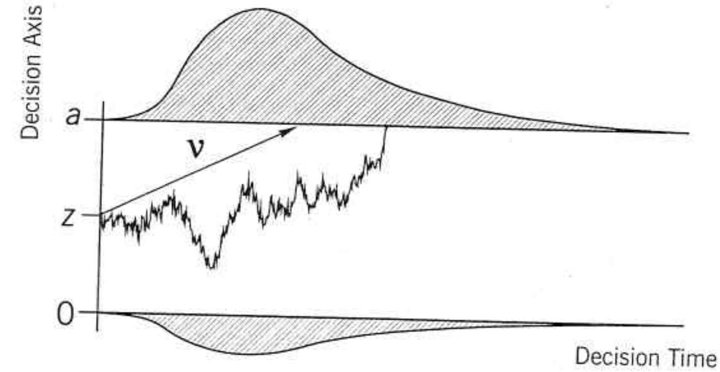

FIGURE 9.4. The Diffusion model. The decision axis is the vertical axis, and the decision time axis is the horizontal axis. The lower threshold is positioned at zero and the upper threshold at *a*. Information accumulation begins at *z* with mean drift rate *v*.

ACKNOWLEDGMENTS

Preparation of this chapter was facilitated by a grant from the National Science Foundation (No. BCS 0820855) to Jeffrey W. Sherman.

NOTE

1. Following Payne, we refer to the two component processes of PD as the automatic (A) and the controlled (C) component. Nevertheless, we reiterate that the extent to which a construct or process (or model component) possesses one or more features of automaticity (or control) is an empirical question. That is, features of automaticity and control must be demonstrated empirically. In the case of the PD model, the characterizations of the A component as automatic and the C component as controlled are well supported by careful research, summarized later.

REFERENCES

Allen, T. J., Sherman, J. W., & Klauer, K. C. (2009). *Social context and the self-regulation of implicit bias.* Unpublished manuscript.

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. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink response and self-report. *Journal of Personality and Social Psychology, 84,* 738–753.

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.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, Jr. & 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.

Bartholow, B. D., Dickter, C. L., & Sestir, M. A. (2006). Stereotype activation and control of race bias: Cognitive control of inhibition and its impairment by alcohol. *Journal of Personality and Social Psychology, 90,* 272–287.

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

Beer, J. S., Stallen, M., Lombardo, M. V., Gonsalkorale, K., Cunningham, W. A., & Sherman, J. W. (2008). The Quadruple Process model approach to examining the neural underpinnings of prejudice. *NeuroImage, 43,* 775–783.

Bishara, A. J., & Payne, B. K. (2009). Multinomial process tree models of control and automaticity in weapon misidentification. *Journal of Experimental Social Psychology, 45,* 524–534.

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

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81,* 828–841.

Brendl, C. M., Markman, A. B., & Messner, C. (2001). How do indirect measures of evaluation work?: Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology, 81,* 760–773.

Brewer, M. B., & Feinstein, A. S. H. (1999). Dual processes in the cognitive representations of persons and social categories. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 255–270). New York: Guilford Press.

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

Chen, S., & Chaiken, S. (1999). The Heuristic Systematic Model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73–96). New York: Guilford Press.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology, 89,* 469–487.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology, 83,* 1314–1329.

Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology, 92,* 1006–1023.

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.

De Houwer, J. (2003). The Extrinsic Affective Simon Task. *Experimental Psychology, 50,* 77–85.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18.

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

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.

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.

Ferguson, M. J., & Bargh, J. A. (2007). Beyond the attitude object: Implicit attitudes spring from object-centered contexts. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 216–246). New York: Guilford Press.

Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 231–254). New York: Guilford Press.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692–731.

Gawronski, B., Deutsch, R., LeBel, E. P., & Peters, K. R. (2008). Response interference as a mechanism underlying implicit measures: Some traps and gaps in the assessment of mental associations with experimental paradigms. *European Journal of Psychological Assessment, 24,* 218–225.

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology, 44,* 1355–1361.

Glaser, J., & Knowles, E. D. (2007). Implicit motivation to control prejudice. *Journal of Experimental Social Psychology, 44,* 164–172.

Gonsalkorale, K., Allen, T. J., Sherman, J. W., & Klauer, K. C. (2009). *Mechanisms of exemplar exposure effects on implicit attitudes.* Unpublished manuscript.

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2009). Aging and prejudice: Diminished regulation of automatic race bias among older adults. *Journal of Experimental Social Psychology, 45,* 410–414.

Gonsalkorale, K., von Hippel, W., Sherman, J. W., & Klauer, K. C. (2009). Bias and regulation of bias in intergroup interactions: Implicit attitudes toward Muslims and interaction quality. *Journal of Experimental Social Psychology, 45,* 161–166.

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

Govorun, O., & Payne, B. K. (2006). Ego depletion and prejudice: Separating automatic and controlled components. *Social Cognition, 24,* 111–136.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

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

Greenwald, A. G., Oakes, M. A., & Hoffman, H. G. (2003). Targets of discrimination: Effects of race on responses to weapons holders. *Journal of Experimental and Social Psychology, 39,* 399–405.

Huntsinger, J. R., Sinclair, S., & Clore, G. (2009). Affective regulation of implicitly measured attitudes and stereotypes: Automatic and controlled processes. *Journal of Experimental Social Psychology, 45,* 560–566.

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., 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.

Klauer, K. C. (2009). Spontaneous evaluations. In F. Strack & J. Förster (Eds.), *Social cognition: The basis of human interaction* (pp. 199–217). New York: Psychology Press.

Klauer, K. C., & Voss, A. (2008). Effects of race on responses and response latencies in the weapon identification task: A test of six models. *Personality and Social Psychology Bulletin, 34,* 1124–1140.

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Processing components of the Implicit Association Test: A diffusion-model analysis. *Journal of Personality and Social Psychology, 93,* 353–368.

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.

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.

Luce, R. D. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Review of Psychology, 46,* 1–26.

Luce, R. D. (1999). Where is mathematical modeling in psychology heading? *Theory and Psychology, 9,* 723–737.

Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin, 107,* 401–413.

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.

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General, 132,* 455–469.

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.

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.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44,* 190–204.

Myung, I. J., & Pitt, M. (2002). Mathematical modeling. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology* (3rd ed., pp. 429–460). New York: Wiley.

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

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6,* 101–115.

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. (2005). Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology, 89,* 488–503.

Payne, B. K. (2008). What mistakes disclose: A Process Dissociation approach to automatic and controlled processes in social psychology. *Social and Personality Psychology Compass, 2,* 1073–1092.

Payne, B. K., Burkley, M., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge?: The role of structural fit. *Journal of Personality and Social Psychology, 94,* 16–31.

Payne, B. K., Lambert, A. J., & Jacoby, L. L. (2002). Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. *Journal of Experimental Social Psychology, 38,* 384–396.

Petty, R. E., & Wegener, D. T. (1999). The Elaboration Likelihood Model: Current status and controversies. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 41–72). New York: Guilford Press.

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

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement. *Journal of Experimental Social Psychology, 44,* 386–396.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

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

Rosano, C., Aizenstein, H., Cochran, J., Saxton, J., De Kosky, S., Newman, A., et al. (2005). Functional neuroimaging indicators of successful executive control in the oldest old. *NeuroImage, 4,* 881–889.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology, 91,* 995–1008.

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.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50.

Stahl, C., & Degner, J. (2007). Assessing automatic activation of valence: A multinomial model of EAST performance. *Experimental Psychology, 54,* 99–112.

Stewart, B. D., von Hippel, W., & Radvansky, G. A. (2009). Age, race, and implicit prejudice: Using process dissociation to separate the underlying components. *Psychological Science, 20,* 164–168.

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

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 & A. Tesser (Eds.), *The construction of social judgment* (pp. 37–65). Hillsdale, NJ: Erlbaum.

Zajonc, R. B. (1965). Social facilitation. *Science, 149,* 269–274.