Title
Separating Multiple Processes in Implicit Social Cognition: The Quad Model of Implicit Task Performance

Journal
Journal of Personality and Social Psychology, 89(4)

ISSN
0022-3514

Authors
Conrey, Frederica R
Sherman, Jeffrey W
Gawronski, Bertram
et al.

Publication Date
2005-10-01

DOI
10.1037/0022-3514.89.4.469

Peer reviewed

Separating Multiple Processes in Implicit Social Cognition: The Quad Model of Implicit Task Performance

Frederica R. Conrey
Indiana University

Jeffrey W. Sherman
University of California, Davis

Bertram Gawronski
University of Western Ontario

Kurt Hugenberg
Miami University

Carla J. Groom
KRC Research

The authors argue that implicit measures of social cognition do not reflect only automatic processes but rather the joint contributions of multiple, qualitatively different processes. The quadruple process model proposed and tested in the present article quantitatively disentangles the influences of 4 distinct processes on implicit task performance: the likelihood that automatic bias is activated by a stimulus; that a correct response can be determined; that automatic bias is overcome; and that, in the absence of other information, a guessing bias drives responses. The stochastic and construct validity of the model is confirmed in 5 studies. The model is shown to provide a more nuanced and detailed understanding of the interplay of multiple processes in implicit task performance, including implicit measures of attitudes, prejudice, and stereotyping.

Keywords: automaticity, implicit measures, multinomial model, process dissociation, controlled processing

In two groundbreaking articles, Schneider and Shiffrin (1977; Shiffrin & Schneider, 1977) deftly organized numerous disparate cognitive processes into two categories: automatic processes and controlled processes. According to their theory, an *automatic process* consists of the spontaneous activation of an existing sequence of nodes in memory. Such processes are effortless and inevitably initiated by the presence of a triggering stimulus. In a *controlled process*, by contrast, a temporary sequence of nodes is established to complete a specific task. Such processes are constrained by cognitive resources but have the advantage of being easily altered, applied, and terminated (Bargh, 1994; see also Sloman, 1996; Smith & DeCoster, 2000; Strack & Deutsch, 2004). The distinction between automatic and controlled processes now occupies a central role in many areas of social psychology and is reflected in contemporary dual-process theories of prejudice and stereotyping (e.g., Devine, 1989), attitude-behavior consistency (e.g., Fazio, 1990), dispositional attribution (e.g., Gilbert, 1989; Trope, 1986), persuasion (e.g., Chen & Chaiken, 1999; Petty & Wegener, 1999), and person perception (e.g., Brewer, 1988; Fiske & Neuberg, 1990).

The most common means of examining the influences of automatic and controlled processes in social psychology has been to administer two separate measures, one aimed at tapping an automatic process and one aimed at tapping a controlled process. For example, most research on prejudice assesses automatic prejudicial responses with implicit measures, such as affective priming (Fazio, Jackson, Dunton, & Williams, 1995) or the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), and conscious, controlled tendencies with explicit measures, such as the Modern Racism Scale (McConahay, 1986) or feeling thermome-

Frederica R. Conrey, Department of Psychology, Indiana University; Jeffrey W. Sherman, Department of Psychology, University of California, Davis; Bertram Gawronski, Department of Psychology, University of Western Ontario; Kurt Hugenberg, Department of Psychology, Miami University; Carla J. Groom, KRC Research, London, United Kingdom.

This research was supported by National Institute of Mental Health Grant MH59774 to Jeffrey Sherman and by Deutsche Forschungsgemeinschaft Grant Ga 747/1-1 to Bertram Gawronski. We thank Rainer Banse, Edgar Erdfelder, Wilhelm Hofmann, Jacob Matthews, Thorsten Meiser, Cindy Pickett, and Jim Sherman for helpful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Frederica Conrey, Department of Psychology, Indiana University, 1101 East Tenth Street, Bloomington, IN 47405, E-mail: rconrey@indiana.edu; to Jeffrey Sherman, Department of Psychology, University of California, One Shields Avenue, Davis, CA 95616, E-mail: jsherman@ucdavis.edu; or to Bertram Gawronski, Department of Psychology, University of Western Ontario, Social Science Centre, London, Ontario N6A 5C2, Canada, E-mail: bgawrons@uwo.ca

Journal of Personality and Social Psychology, 2005, Vol. 89, No. 4, 469–487

0022-3514/05/$12.00 DOI: 10.1037/0022-3514.89.4.469

ters (Haddock, Zanna, & Esses, 1993). This approach has led social psychologists to largely equate implicit measures with automatic processing and explicit measures with controlled processing.

Although this task dissociation approach has been responsible for many significant advances in social psychology, it has certain limitations. First, it confounds processing style (automatic vs. controlled) with the particular measurement task. As such, the tasks may differ in a number of ways beyond the extent to which they tap automatic versus controlled processes. For example, many observed dissociations between implicit and explicit memory tasks may be reinterpreted as dissociations between tasks that tap perceptual versus conceptual processes (e.g., Roediger, 1990; Sherman, Lee, Bessenoff, & Frost, 1998).

The more general point is that no task is "process pure." It is technically impossible that any task that requires observable responses depends entirely on automatic processes and not at all on controlled processes. Moreover, it is quite unlikely that any task depends entirely on controlled processes and not at all on automatic processes. Rather, most, if not all, of the behaviors researchers wish to understand will be influenced by simultaneously occurring automatic and controlled processes that influence one another (Wegner & Bargh, 1998). Thus, attempts to isolate particular processing styles with separate tasks will be incapable of identifying the complexity of automatic and controlled influences in producing discrete responses and will necessarily oversimplify conclusions about behavior.

Multiple Automatic and Controlled Processes

Another important issue in the context of automatic and controlled processes is the question of qualitative differences between processes. Although the distinction between automatic and controlled processing is ubiquitous in social psychology, different formulations of this distinction emphasize different individual processes.

Conceptualizations of Control

Control in dual-process theories (cf. Chaiken & Trope, 1999) is most commonly understood as acting to distill information or to determine a correct answer. In dual-process models of persuasion (Chen & Chaiken, 1999; Petty & Wegener, 1999), for example, control is exerted in weighing the strengths and weaknesses of a persuasive message. In a similar vein, dual-process models of person perception argue that forming an accurate impression requires controlled processing of individuating in contrast to category information (e.g., Brewer, 1988; Fiske & Neuberg, 1990).

Recently, another important type of controlled process, self-regulation, has received increasing attention. Wegner's (1994) model of thought suppression contends that when people attempt to suppress specific thoughts, such as thoughts of a white bear, two processes are engaged: an automatic monitoring process that scans memory for thoughts of white bears and a controlled operating process that suppresses those thoughts when they are discovered. Such controlled regulatory efforts play an important role in dual-process models of prejudice and stereotyping, proposing that effortful control is necessary to overcome automatically activated stereotypes (e.g., Devine, 1989).

Historically, dual-process theories focus on only one of these two roles of control. However, though they may be similar in that they both require cognitive resources, it is clear that accuracy assessment and self-regulation are very different and that both processes may operate simultaneously in many contexts, often with very different results. For example, a police officer's decision about whether or not to shoot a Black man who may or may not have a gun depends both on his ability to discriminate whether or not the man has a gun and, if he has no gun, his ability to overcome an automatic bias to associate Blacks with guns and to shoot (cf. Correll, Park, Judd, & Wittenbrink, 2002; Greenwald, Oakes, & Hoffman, 2003; Payne, 2001). An accurate depiction of complex behavior must consider both processes simultaneously.

Conceptualizations of Automaticity

The role of automatic processes in determining responses also has been conceptualized in two different ways. The first was described by Schneider and Shiffrin (1977) as the spontaneous activation of existing associations that capture attention and draw it away from deliberate cognition toward the activated sequence. In his work on affective primacy, for instance, Zajonc (1980) showed that objects are processed affectively before any controlled processing is engaged. Later work showed that this automatic activation of affective associations can interfere with deliberate responding (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Such interference effects form the basis for modern implicit measures of attitudes such as affective priming (Fazio et al., 1995) or the IAT (Greenwald et al., 1998).

In other tasks, however, the role of automatic bias has been understood differently. Memory research typically focuses on the role of bias in facilitating responses when control fails. In proposing his process dissociation procedure, which is discussed in more detail below, Jacoby (1991) pointed out that either controlled memory search or an automatic feeling of familiarity could lead to the correct identification of old items on a memory test. This response bias is qualitatively different from the automatic activation of associations. Rather than interfering with controlled responding, response bias influences the response only when control fails. The exact nature of this form of bias might be anything from Jacoby's (1991) familiarity bias to the surprisingly powerful bias to prefer items placed on the right side of a display (Nisbett & Wilson, 1977).

Again, dual-process models have typically focused on one or the other type of bias, either automatic association activation or response bias, but not both. Yet, here too, it is clear that many responses may be influenced simultaneously by both processes. A police officer's split-second decision to pull the trigger in response to a Black man who may be pointing a gun at him might be influenced by automatically activated associations between Black men and aggression (e.g., Correll et al., 2002; Greenwald, Oakes, & Hoffman, 2003; Payne, 2001). In the absence of such associations, however, the officer's decision still might be influenced by an implicit bias to presume that he is in danger in the absence of clear evidence to the contrary.

Multiple Processes in Implicit Measures

The purpose of the quadruple process model (quad model) proposed in this article is to estimate the simultaneous contributions of both types of processes that have typically been labeled *automatic* as well as both types of processes that have usually been labeled *controlled*. Specifically, we contend that responses on implicit measures depend on the automatic activation of an association (*association activation*), the ability to determine a correct response (*discriminability*), the success at overcoming automatically activated associations (*overcoming bias*), and the influence of any response bias that may influence overt reactions in the absence of other available guides to response (*guessing*).

An illustrative example of the joint contribution of these four processes is the IAT, developed by Greenwald et al. (1998). The IAT is a double discrimination task in which participants are asked to simultaneously categorize target stimuli (e.g., Black and White faces) and attribute stimuli (e.g., pleasant and unpleasant words). For example, in the compatible block of an IAT designed to assess White participants' automatic preference for Whites over Blacks, participants are asked to respond to pleasant words and White faces with one key and to unpleasant words and Black faces with another key. On the incompatible block, the response pairings are switched (i.e., Black–pleasant, White–unpleasant). To the extent that judgments on the second block are more difficult than categorizations on the first, participants are thought to have an automatic preference for Whites over Blacks.

Notwithstanding the successful use of the IAT in various areas, it seems likely that the observable responses required by the IAT are not determined exclusively by automatic associations (see Brendl, Markman, & Messner, 2001; McFarland & Crouch, 2002; Mierke & Klauer, 2003; Rothermund & Wentura, 2004). In a Black–White IAT, for example, association activation may be responsible for an automatic tendency to respond "negative" to a Black face (association activation). Depending on the particular key assignment, this automatic tendency may be congruent or incongruent with the correct answer "Black" achieved through discrimination (discriminability). If the task requires pairing Blacks with negative words, then the responses provided by automatic associations and discrimination are compatible. In this case, there is no conflict, and there is no need to overcome bias in order to produce the correct response. However, if the two response tendencies are incongruent (pairing Blacks with positive words), then whether the automatic associations or accurate discrimination finally drives the response is determined by whether the participant succeeds in overcoming his or her associations (overcoming bias). Finally, if no association is activated and the correct response is not available, then participants must guess (guessing). In this case, participants may guess right or left randomly. However, participants may also exhibit an unintentional tendency to favor the right-hand side of a display (Nisbett & Wilson, 1977) or even a strategic tendency to respond with the positive key in order to avoid looking prejudiced.

Though we have used the IAT as an example, the present considerations can be applied to any kind of implicit measure that is based on the logic of response compatibility (cf. De Houwer, 2003; Kornblum, Hasbroucq, & Osman, 1990), including affective priming (Fazio et al., 1995), the Stroop task (Kawakami, Dion, & Dovidio, 1999), the go/no-go association task (Nosek & Banaji, 2001), and other sequential priming tasks that rely on processes of response compatibility (Payne, 2001).[1] All of these tasks manipulate whether an automatic association is congruent or incongruent with a correctly discriminated response.

Standard techniques used to analyze data from implicit measures cannot disentangle the contributions of these four processes (association activation, discriminability, overcoming bias, and guessing) that may influence responses on implicit tasks. For example, these tasks cannot distinguish between people who have strong automatic associations that they are able to overcome from people who have weak associations. However, given the importance attributed to the interplay of automatic and controlled processes in social psychology (Chaiken & Trope, 1999; Smith & DeCoster, 2000; Strack & Deutsch, 2004), it seems highly desirable to have a methodological tool that is able to isolate these processes. The quad model provides a means of statistically estimating the values of the four processes from observed error rates.

Measuring Multiple Processes: Process Dissociation

The quad model proposed and tested in the present article is substantially influenced by Jacoby's work on process dissociation (e.g., Jacoby, 1991; Jacoby, McElree, & Trainham, 1999; Lindsay & Jacoby, 1994). For this reason, we first illustrate the general idea of process dissociation by discussing the two major models of process dissociation and then outline the basic assumptions of the quad model.

The "C-First" Model of Process Dissociation

Jacoby's C-first model of process dissociation (Jacoby, 1991) focuses on the role of accurate discrimination and response bias. This model was developed to disentangle the contributions of controlled recollection and automatic familiarity to recognition memory. It relies on contrasting two types of trials: compatible trials, on which recollection and familiarity should lead to the same response, and incompatible trials, on which the two processes should lead to different responses. To the extent that familiarity determines responses, performance on incompatible trials will be poor compared with that on compatible trials.

In one of the first studies using the procedure (Jacoby, 1991), participants studied two lists of words. For the sake of simplicity, we refer to the two lists as the *red list* and the *blue list*. In a first condition, participants had to distinguish between old words from the two lists and new foil words that were not part of the lists (standard recognition task). In this task, items from both the red and the blue list could be correctly classified as old on the basis of either the participants' ability to consciously recollect having seen the items or the feeling of familiarity evoked by the items. As such, recollection and familiarity work in concert for both the red and the blue list.

In a second condition, participants were instructed to respond "old" only to words from the red list. Words from the blue list, as well as new foil items, were to be labeled *new* (modified recognition task). When participants have explicit recollection memory about whether the word was part of the red or the blue list, words from the red list will correctly be judged as old, and words from the blue list will correctly be judged as new. However, when participants have no explicit recollection memory, they may rely on the familiarity of the word to make their "old versus new" judgment. In this case, words from both the red and the blue list will be judged old, resulting in correct judgments for items from the red list but in incorrect judgments for items from the blue list. In other words, recollection and familiarity still work in concert for

[1] Note that another prominent implicit measure, Wittenbrink, Judd, and Park's (1997) semantic priming task, is not based on processes of response compatibility, and thus cannot be analyzed with the model proposed in the present article (cf. De Houwer, 2003). A discussion of important differences between semantic priming and response compatibility tasks can be found in Gawronski and Bodenhausen (2004).

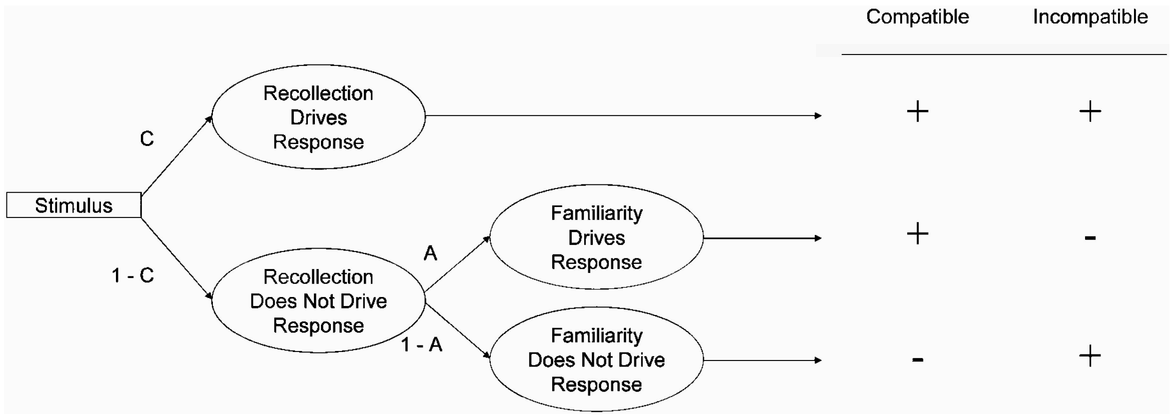

Figure 1. The C-first model (Jacoby, 1991). Each path represents a likelihood. Parameters with lines leading to them are conditional upon all preceding parameters. The table on the left panel of the figure depicts correct (+) and incorrect (-) responses for the standard recognition task (compatible) and the modified recognition task (incompatible).

items from the red list, but they work against one another for items from the blue list.

Responses in this task can be depicted in a processing tree (see Figure 1). In the tree, each path represents a likelihood. C represents the likelihood that controlled recollection will determine the response. For example, if recollection succeeds, then the answer to items from the blue list will be correct for both the standard recognition task (compatible) and for the modified recognition task (incompatible). If recollection fails (1–C), however, then feelings of familiarity (A) may drive the response. If this happens, then items from the blue list will be correctly labeled *old* in the standard recognition task (compatible). However, they will be incorrectly labeled *old* in the modified recognition task (incompatible). Finally, if a blue item cannot be recollected (1–C) and participants base their judgments on the nonfamiliarity of the word (1 – A), then items from the blue list will be incorrectly labeled *new* in the standard recognition task (compatible), but they will be correctly labeled *new* in the modified recognition task (incompatible).[2] Using the observed error rates in the four conditions, it is possible to solve for A and C algebraically.

This model of process dissociation has been successfully applied across a number of domains (e.g., Jacoby, 1991; Payne, 2001; Payne, Lambert, & Jacoby, 2002; Sherman, Groom, Ehrenberg, & Klauer, 2003; Toth, Reingold, & Jacoby, 1994). However, because it limits the role of bias to the case in which recollection or control fails, the C-first model is appropriate for modeling response bias but not for modeling tasks in which automatic associations attempt to capture the response even though the correct response is available. For example, there is no way for the model to account for the Stroop color-naming task (Stroop, 1935) in which most people can determine the color of the ink easily, but the automatic habit to read the word draws the response away from the correct answer. In a similar vein, the model does not account for the IAT in which most people can accurately discriminate the presented stimuli (e.g., Black and White faces), but automatic associations may interfere with a correct response.

The "A-First" Model of Process Dissociation

To address this limitation, Lindsay and Jacoby (1994) proposed a second model of process dissociation. The A-first model also estimates two parameters that have been generalized here to the automatic component, A, and the controlled component, C, for the sake of consistency. An illustrative application of the A-first model is the Stroop color-naming task (Stroop, 1935). When the word and the ink color are incompatible (e.g., the word "*Red*" printed in blue ink), the correct response is typically much harder to provide than when the color and the ink are compatible (e.g., the word *Red* printed in red ink). This interference effect can be conceptualized in terms of the A-first model (see Figure 2). When the automatic habit to read the word drives the response (A), the correct answer will be given when the word and the ink color are compatible, and the incorrect answer will be given when the word and the ink color are incompatible. If, however, the automatic habit does not drive the response (1 – A), then explicit knowledge of the ink color can drive the response (C), providing the correct answer regardless of compatibility. Finally, if control does not drive the response (1 – C), then the model assumes that the incorrect answer will be returned.

Even though the A-first model, in contrast to the C-first model, is generally appropriate for tasks in which automatic associations attempt to capture the response, it still has limitations. Specifically, the A parameter in the A-first model estimates both the joint probability that the bias is activated and that it drives the response. Thus, the A parameter does not account for cases in which self-regulation succeeds. With regard to the Stroop task, for example, the A-first model would not be able to differentiate between an illiterate child who has no word-reading habit and a highly motivated adult who succeeds in overcoming the habit consistently. Distinguishing cases in which an automatic response is not activated from cases in which the response is activated but successfully inhibited has become a critical question in research on prejudice. In this research, demonstrations of diminished prejudice on implicit measures may be interpreted as reflecting attitude change (e.g., Blair, Ma, & Lenton, 2001; Dasgupta & Greenwald, 2001; Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000; Rudman, Ashmore, & Gary, 2001) or the enhanced ability to overcome

[2] Note that if the judgment is not driven by explicit recollection memory (1 – C) and a familiarity-consistent judgment is not made (1 – A), the model assumes that a familiarity-inconsistent judgment will be made.

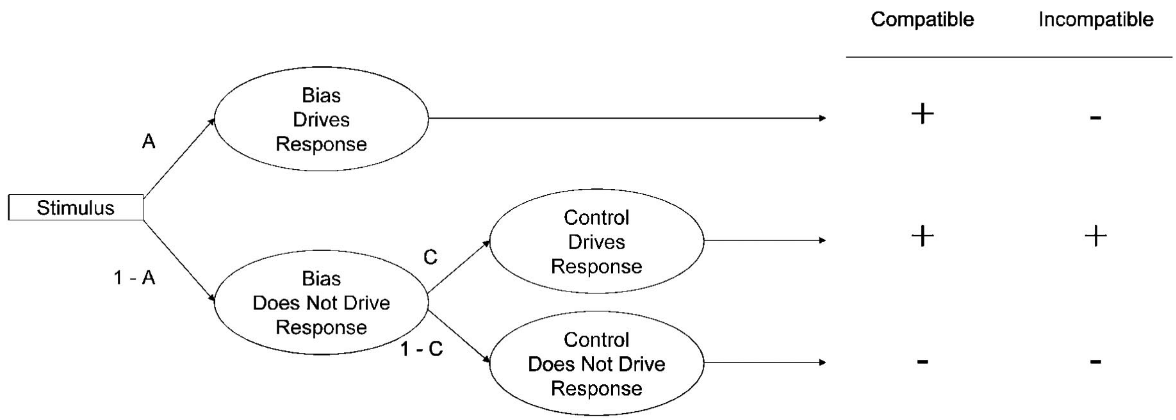

Figure 2. The A-first model (Lindsay & Jacoby, 1994). Each path represents a likelihood. Parameters with lines leading to them are conditional upon all preceding parameters. The table on the left panel of the figure depicts correct (+) and incorrect (-) responses for compatible (e.g., the word *red* printed in red ink) and incompatible items (e.g., the word *red* printed in blue ink) in a Stroop task.

bias rather than attitude change, per se (e.g., Devine & Monteith, 1999; McFarland & Crouch, 2002; Moskowitz, Gollwitzer, Wasel, & Schaal, 1999). Another limitation of the A-first model is that when there is no automatically activated habit, and the correct response cannot be determined through deliberation, the assumption when working with this model is that an incorrect response will be given. There is no way of using this model to account for guessing that may occasionally return a correct response (Buchner, Erdfelder, & Vaterrodt-Pluennecke, 1995).

The Quad Model

The quad model (see Figure 3) is a multinomial model (Batchelder & Riefer, 1999; Riefer & Batchelder, 1988; for a discussion of the range and limits of multinomial models in social psychology, see Klauer & Wegener, 1998) designed to disentangle four qualitatively distinct processes that contribute to overt responses in implicit measures on the basis of the logic of response compatibility: the automatic activation of an association (*association activation*; AC), the ability to determine a correct response (*discriminability*; D), the success at overcoming automatically activated associations (*overcoming bias*; OB), and the influence of a general response bias that may guide responses in the absence of other available guides to response (*guessing*; G). In the tree, each path represents a likelihood. Parameters with lines leading to them are conditional upon all preceding parameters. For instance, OB is conditional upon both AC and D. In a similar vein, G is conditional upon no AC (1 – AC) and no D (1 – D).[3]

The AC Parameter

AC, the association activation parameter, reflects the likelihood that an association is automatically activated by a stimulus. The opposite probability, 1 – AC, represents the likelihood that the association is not activated. The AC parameter can be understood as the strength of the association activated by the stimulus. The stronger the association, the more likely it will be activated by a relevant stimulus. The AC parameter directly reflects what implicit measures of social cognition are typically used to assess.

The D Parameter

D estimates discriminability. The controlled process of discrimination corresponds to the most typical role of control in dual-process theories, the application of effort in determining the correct response. For instance, in the context of person perception, discrimination would determine individuation in the face of an automatic tendency to generalize on the basis of group membership (Brewer, 1988; Fiske & Neuberg, 1990). It is important to emphasize that D represents the likelihood that the answer can be determined rather than the likelihood that the answer is determined. D is knowledge-based and thus sensitive to the availability of relevant information in memory. Moreover, the use of D also includes being sensitive to the amount of attention paid to the stimulus and to cognitive capacity. Thus, D should be lower if a person is distracted, engaged in worrying about the task, or in counting the ceiling tiles. Finally, the use of D should also include being sensitive to motivation. Greater motivation to succeed on the task should lead to greater allocation of resources, and thus to a higher D.[4]

[3] Though some parameters are conditional on others, the processing tree does not necessarily imply a temporal sequence of processes. The likelihood represented by the AC parameter, for instance, can tell us whether an association was activated or not, but not whether it was activated before or after the correct response (D) was determined.

[4] In some tasks, it is possible that the discriminability of the stimulus differs, depending on whether an automatic association is activated. For example, people with stronger spontaneous negative reactions to snakes may be better able to discriminate whether the object in the grass is a snake or a stick than people with weaker negative reactions. Though we acknowledge that there are contexts in which D might differ across AC and 1 – AC cases, for the sake of parsimony and stringency of our tests of the model, we have set them equal in the current research. In disputable cases, this assumption can be empirically tested, as long as the model remains locally identifiable within the context of the task. In all of the present research, the quad model fits well with this criterion in place.

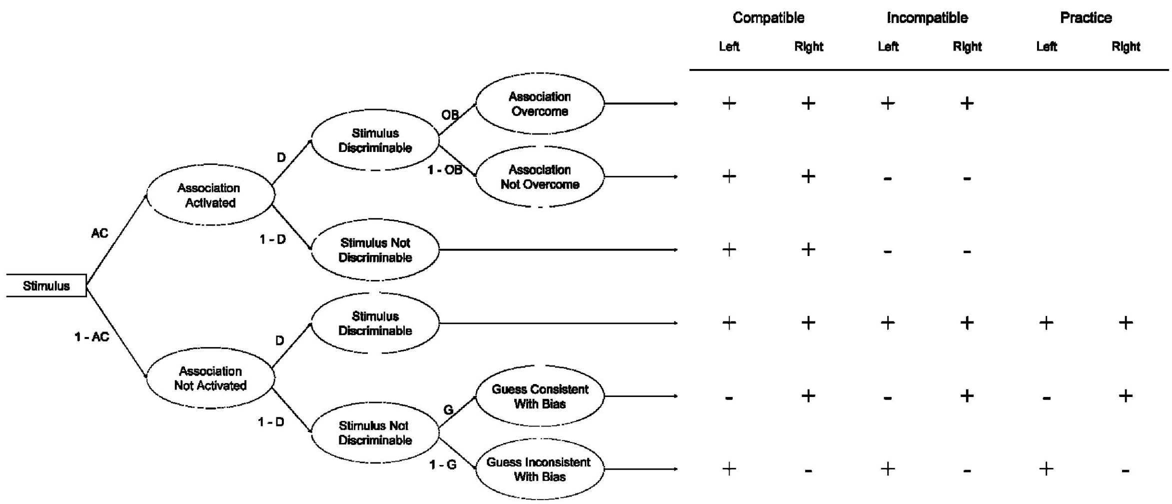

Figure 3. The quadruple process model (quad model). Each path represents a likelihood. Parameters with lines leading to them are conditional upon all preceding parameters. The table on the left panel of the figure depicts correct (+) and incorrect (-) responses in the Implicit Association Test as a function of process pattern and block type.

The OB Parameter

Just as interesting as the case in which automatically activated associations do drive the response is the case in which associations are activated but overcome in favor of deliberate responding. Note, however, that OB represents a different kind of use of control than discrimination. Rather than representing control exerted in the service of individuation, OB represents control exerted in the service of inhibition. The OB parameter reflects success at overcoming bias. When bias is activated (AC) and there is explicit information in the environment or in memory that could be used to make a deliberate judgment (D), associative and rule-based processing (Smith & DeCoster, 2000) can be seen as competing to drive the response, particularly if the two sources provide incompatible information. OB moderates between these two processes. If the bias is overcome (OB), then discrimination (D) drives the response. However, if the bias is not overcome (1 – OB), then the automatic bias (AC) drives the response. The estimated OB is the probability that an activated bias is overcome in favor of a deliberate response. Because OB represents a controlled process, it, like D, should be influenced by both cognitive capacity and motivation.

The G Parameter

When no association is activated and there is no correct answer available, a guess must be made. The G parameter represents a general response bias like the bias component in signal detection theory (Green & Swets, 1966). Though G does represent the influence of a bias, it does not necessarily reflect a purely automatic process. In the IAT, for example, G can reflect the impact of an unconscious tendency to respond with the right hand, but it can also reflect a strategic bias to respond with the positive key, given that it may appear less prejudiced to incorrectly assign a Black face to the "positive" side of the screen. In contrast to AC, which represents the likelihood that an automatic association will be activated, G represents the likelihood that a response bias is activated and drives the response.

Analyzing Data With the Quad Model

In contrast to most data analytic strategies for implicit measures, error rates rather than response latencies are used in the quad model. Specifically, individual parameter likelihoods are estimated in the quad model for the four processes from the observed probability of a correct response given a particular stimulus type (e.g., a White face in the incompatible block of the IAT). Multinomial models, such as the quad model, are fit to data by matching error rates predicted by the parameters to the observed error rates in the sample. Each of the paths from left to right in the processing tree represented in Figure 1 represents a compound probability (e.g., AC $\times$ D $\times$ OB) and predicts a specific response (i.e., correct or incorrect). The sum of all the probabilities associated with a response is the total probability of that response (for a general introduction to multinomial modeling, see Klauer & Wegener, 1998).

For instance, the model predicts that a White face in the incompatible block of a Black–White IAT will be assigned to the correct side of the screen with the probability: $p(\text{correct} \mid \text{White, incompatible}) = AC \times D \times OB + (1 - AC) \times D + (1 - AC) \times (1 - D) \times (1 - G)$. 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 is activated and that the correct answer can be discriminated and that the association is overcome in favor of controlled responding. 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 determined. Finally, $(1 - AC) \times (1 - D) \times (1 - G)$ is the likelihood that the association is not activated and the correct answer cannot be discriminated and that the participant guesses the left-hand response. For a White face, this guess would return the correct response. The sum of these probabilities is the total probability of a correct response for the item (for more details, see Appendix A).

Parameter values are estimated by creating equations for correct versus incorrect responses for each item type. A chi-square value is then computed using the observed error rate and the expected error rate provided by the model. Obviously, the smaller this value, the better, so the parameters are changed through maximum-likelihood estimation (MLE) until they return a minimum possible value of the chi-square. The parameter values resulting from this procedure can then be interpreted as the (relative) level of the corresponding process. If the chi-square is not significant, then the model is said to fit the data (for more details, see Appendix A).

Of as much interest as the model fit is the specific pattern of the parameters in the model. In order to establish the validity of the parameters (i.e., Do they represent what we say they do?), they must be shown to vary independently of one another, and they must be shown to respond appropriately to key manipulations. In testing such hypotheses within a multinomial model, two or more parameters are set equal to each other or to a given value (in our case, 0 or 0.5). If the model fit is significantly diminished when two parameters are set equal, then the two parameters are significantly different from one another and cannot be combined. If the fit is significantly diminished by setting a parameter to 0, then that parameter is significantly greater than 0 and must be included in the model (see Appendix B).

Testing the Quad Model

To test the quad model, we applied it to the IAT, developed by Greenwald et al. (1998), and a sequential priming paradigm, developed by Payne (2001). These tasks are designed to measure automatic associations between categories of stimuli. In conducting the following research, we had three general goals. First, we wanted to establish the stochastic and construct validity of the model and its parameters. Second, we wanted to demonstrate the quad model's usefulness as a tool for examining effects on automatic and controlled processes that cannot be as well illuminated by conventional data analytic procedures. Finally, we wanted to use the model to provide a better understanding of the simultaneous influence of automatic and controlled processes on implicit measures.

In Study 1, we fit the model to data from an IAT assessing automatic evaluations of flowers and insects, showing that performance on the task is influenced by multiple processes and that the quad model's parameters vary in meaningful ways. In Study 2, we fit the model to data from an IAT assessing automatic evaluations of Black and White people. This study further validates the model, showing that parameters estimating controlled processes (D and OB) vary as a function of processing constraints, whereas the AC parameter, reflecting automatic association activation, does not. Study 3 was designed to test the validity of the G parameter, showing that asymmetric base rates in the types of required responses systematically influence the G parameter, but leave the other parameters (AC, D, OB) unaffected. In Study 4, we turn from validating the model to using the parameters to describe and predict data revealed by standard data analytic strategies. Individual parameter estimates from a Black–White IAT were used to predict conventional IAT latency scores of implicit prejudice. Consistent with the interpretation of the parameters, the pattern of prediction shows a positive relation between conventional IAT latency scores and AC, but a negative relation to success at OB. Finally, in Study 5, we reanalyzed data obtained in a sequential priming task designed to assess automatic associations between Blacks/Whites and guns (Lambert et al., 2003). This analysis shows that results obtained with a less complex model of process dissociation mask important effects of a public versus private manipulation on the automatic activation of associations.

Study 1

The primary goal of Study 1 was to apply the quad model to a standard flowers-insects IAT, using photographs of flowers and insects as target items and positive and negative words as attribute items. In fitting the model, we allowed the parameters to vary across theoretically meaningful dimensions but set them equal across dimensions on which they should not vary. First, we estimated two AC parameters, one for each of the associations tapped by the task, flowers–pleasant and insects–unpleasant. The dissociation is based on the assumption that an individual's automatic associations related to flowers can be independent from his or her automatic associations related to insects. Second, we allowed OB to vary across target items and attribute words. This was done because only target items may trigger antagonistic responses in terms of AC and D (e.g., a picture of a spider triggers the response "unpleasant" for AC and the response "insect" for D), whereas AC and D for attribute items usually lead to the same response (e.g., the word *hate* triggers the response "unpleasant" for both AC or D). Hence, participants may spend more effort overcoming bias on target than on attribute items. Finally, we estimated a single D parameter, estimating the ability to accurately categorize the presented stimuli, and a single G parameter, estimating the tendency to choose the right-hand response when no other information is available.

Method

Twenty-nine undergraduates participated in exchange for partial course credit. All participants completed a standard flowers–insects IAT (Greenwald et al., 1998, Experiment 1). Stimuli for the task were 10 photographs (i.e., 5 insects, 5 flowers) and 10 words (i.e., 5 pleasant words, 5 unpleasant words). The first two blocks were 20-trial practice blocks in which participants first practiced assigning only unpleasant and pleasant words to the right- and left-hand categories and then practiced assigning insects and flowers to the right- and left-hand categories. After the practice blocks, participants were instructed to press the right-hand key for pleasant words and pictures of flowers and the left-hand key for unpleasant words and pictures of insects in a 40-trial compatible test block. This block was followed by another 20-trial practice block in which participants practiced assigning flowers and insects to the opposite sides of the screen. Finally, participants completed a 40-trial incompatible test block, with pleasant words and pictures of insects assigned to the right-hand key and unpleasant words and pictures of flowers assigned to the left-hand key.

Results

Before analyzing the four parameter estimates, we tested whether the quad model fit the data. Consistent with the assumption that the quad model can be used to describe data obtained from an IAT, the model fit the data sufficiently well, $\chi^2(2) = 1.74$, $p = .42$, with an overall error rate of approximately 7%.

The parameter estimates are printed in Table 1. First, we tested whether D differed significantly from zero. Consistent with the assumption that participants are generally able to accurately cate-

Table 1
Parameter Estimates for Flower–Insects IAT, Study 1

Parameter	Type	Estimate
AC	Insect–unpleasant	0.13
	Flower–pleasant	0.38
OB	Attribute judgment	0.77
	Category judgment	0.87
D		0.88
G		0.50

Note. Goodness of model fit: χ^2 (2) = 1.74, p = .42. IAT = Implicit Association Test; AC = association activation; OB = overcoming bias; D = discriminability; G = guessing.

gorize the stimuli presented in the IAT, D was considerably high and significantly differed from zero, $\chi^2(1) = 3005.62$, $p < .001$.

With respect to the G parameter, we tested whether participants exhibited a systematic response bias. In the present application, G represents a right-hand guessing bias. A value of .5 would suggest that guessing was completely random. The parameter was coded so that values higher than .5 indicated a preference for the right-hand category, and values below .5 indicated a preference for the left-hand category. In this study, the G parameter did not differ from .5, $\chi^2(1) = 0.00$, $p = 1.00$.

Our primary prediction for AC was that it would be greater than zero, indicating a significant association activation component in the IAT. That is, participants were expected to show an automatic preference for flowers over insects. This prediction was borne out; the AC parameter for the insect–unpleasant association was greater than zero, $\chi^2(1) = 7.78$, $p = .01$, as was the AC parameter for the flower–pleasant association, $\chi^2(1) = 58.32$, $p < .001$. These results indicate that association activation does indeed play a significant role in determining responses on the IAT. AC for the flower–pleasant association was higher than was AC for the insect–unpleasant association, $\chi^2(1) = 9.47$, $p = .002$, indicating that the IAT is indeed tapping two distinct automatic attitudes.

Our primary prediction for OB was that it, too, would differ significantly from zero, indicating that overcoming bias plays a role in determining responses on the IAT. Consistent with this assumption, the OB parameter for the target items significantly differed from zero, $\chi^2(1) = 4.57$, $p = .03$. Though the OB parameter for the attribute items was only marginally different from zero, $\chi^2(1) = 2.39$, $p = .12$, the two parameters did not differ from each other, $\chi^2(1) = 1.71$, $p = .19$.

Discussion

The present results offer first evidence for the validity of the quad model. The model fit the data from the flowers–insects IAT very well. Moreover, the specific parameter estimates indicated that both AC and OB played important roles in determining responses on the IAT. As expected, D of the stimuli was considerably high. There was no evidence for a general G bias. Supporting the construct validity of AC, the parameter varied with the association—the positive–flower association was stronger than the negative–insect association—but not with the item type. This finding is also consistent with the notion that the IAT measures two distinct associations.

Study 2

Obviously, we cannot conclude that the parameters actually measure what we intend them to measure just because the parameters accurately estimate the observed data. To be valid, the parameters in the model must meet two criteria: (a) They must be able to vary independently of each other and (b) they must vary meaningfully. The aim of Studies 2a and 2b was to show that the parameters that represent controlled processes vary when participants' ability to use control is limited. Whereas automatic processes, such as AC, should be unaffected by participants' restricted ability to engage in controlled processing, we expected a substantial influence of this restriction on more effortful processes such as D and OB.

In order to test these assumptions, a standard IAT was used in Study 2a, similar to that used in Study 1, whereas a response window IAT was used in Study 2b, imposing a time limit on participants' categorization judgments (Cunningham, Preacher, & Banaji, 2001). Such time constraints can be assumed to limit participants' ability to engage in controlled processing, which should reduce parameters reflecting controlled processes but not parameters reflecting automatic processes. Specifically, we expected that AC would vary as a function of the specific attitude. However, AC should not vary as a function of the imposed time constraints. D and OB, however, were expected to vary as a function of the response window manipulation; items should be more discriminable the longer they are on the screen, and participants' ability to overcome automatic bias should be higher when they have more time to respond.

Method

Seventeen students participated in Study 2a, using a standard IAT, and 30 students participated in Study 2b, using a response window IAT. Participants in both studies received course credit for participation. The IATs were variants of a Black–White IAT designed to assess implicit preference for Whites over Blacks (e.g., Greenwald et al., 1998, Experiment 3). Stimuli for the IATs consisted of 10 each of pleasant words, unpleasant words, "Black" names, and "White" names taken from Greenwald et al. (1998). The standard IAT had the same structure as that used in Study 1. The first two blocks were 20-trial practice blocks in which participants first had to categorize pleasant and unpleasant words and then White and Black names. After these practice blocks, the 40-trial compatible test block appeared, and participants responded to pleasant words and White names with the right-hand key and to unpleasant words and Black names with the left-hand key. This block was followed by another 20-trial practice block requiring categorizations of Black and White names with a switch of the response keys. Finally, participants completed a 40-trial incompatible test block, with pleasant words and Black names on the right and unpleasant words and White names on the left.

The same stimuli and structure of the standard IAT was used in the response window IAT. The response window in the IAT began 225 ms after the stimulus appeared and was 450 ms long. Participants were instructed to respond to the stimulus "before time runs out." If participants responded within the time window, then the word turned red. If participants failed to respond before the time window ended, then the word disappeared. A new trial did not begin until participants provided a response.

Results

Overall, the IAT error rate was 6% in the no-window condition and 14% in the response window condition. Because we did not expect the response window manipulation to affect the automatic activation of associations, the AC parameter was set equal across the two conditions. In order to link the processing trees of the two conditions, we estimated G for the test blocks, but not for the practice blocks, to vary separately across the two conditions, although we did not expect any differences in this parameter (see Appendix B). The quad model fit the data with these constraints in place, $\chi^2(5) = 3.19$, $p = .67$.

Parameter estimates for the two IATs are presented in Table 2. As predicted, the D parameter was considerably higher for the standard IAT than for the response window IAT, $\chi^2(1) = 46.47$, $p < .001$. This result is consistent with the assumption that D reflects a controlled process that can be undermined when ability to engage in controlled processing is limited.

Consistent with the results of Study 1, OB was again higher than zero in the no-window condition, $\chi^2(2) = 40.97$, $p < .001$. However, OB did not significantly differ from zero in the response window condition, $\chi^2(2) = 1.09$, $p = .58$. The difference between window and no-window conditions was statistically significant, $\chi^2(2) = 21.15$, $p < .001$. This finding is consistent with the assumption that OB measures a controlled process that is undermined when the ability to engage in cognitive control is depleted. The two OB parameters for target and attribute items did not differ significantly in the no-window condition, $\chi^2(1) = 0.02$, $p = .99$, or in the response window condition, $\chi^2(1) = 1.09$, $p = .30$.

As noted above, the model fit with the AC parameters set equal across the window and no-window conditions. Although we prefer this more stringent, constrained model for these tests of validity, we did test whether allowing the AC parameters to vary would significantly improve the model fit. This was not the case, $\chi^2(2) = 1.79$, $p = .41$. The White–pleasant association was significantly higher than zero, $\chi^2(1) = 100.31$, $p < .001$, as was the Black–unpleasant association, $\chi^2(1) = 40.15$, $p < .001$, indicating a significant association activation component in the IAT. AC for White–pleasant combinations was somewhat higher than AC for Black–unpleasant combinations, $\chi^2(1) = 4.30$, $p = .04$. This finding is consistent with the often observed tendency for in-group favoritism to be stronger than out-group derogation (Brewer, 1999).

G did not differ across the test blocks, $\chi^2(1) = 1.20$, $p = .27$; but it was significantly greater than .5, $\chi^2(2) = 15.73$, $p < .001$, indicating that a right-hand guessing bias drove responses when no other information was available. Of most interest, the G parameter for the practice blocks did not differ from .5, $\chi^2(1) = 0.56$, $p = .45$. This result suggests that the guessing bias in the test blocks may be a strategic bias to guess the "positive" side of the screen (which, in this case, is the right-hand side) rather than a general right-hand bias.

Table 2
Parameter Estimates for Black–White IATs, Study 2

Parameter	Comparison	Estimate	
AC	Black–unpleasant	0.11	
	White–pleasant	0.18	
G	Practice	0.52	
		IAT	RW-IAT
OB	Attributes	0.82	0.00
	Names	0.85	0.22
D		0.89	0.78
G	Test blocks	0.56	0.64

Note. Goodness of model fit: $\chi^2(5) = 3.19$, $p = .67$. IAT = Implicit Association Test; AC = association activation; OB = overcoming bias; D = discriminability; G = guessing.

Discussion

Results from Studies 2a and 2b offer further support for the validity of the quad model. Most important, the present findings indicate that the parameters estimated by the quad model reflect distinct processes that differ in the extent to which they are automatic or controlled. Consistent with an interpretation of AC as the likelihood of automatic association activation, this parameter did not vary as a function of time constraints but did vary as a function of the attitude measured. We predicted and found an effect of time constraints on D. These results are consistent with our interpretation of D as a controlled process. Moreover, we predicted that OB would vary as a function of time constraints. In fact, when the time to respond was restricted, OB dropped significantly. Finally, a right-hand guessing bias emerged on the test blocks in both conditions. This bias may reflect a guessing strategy that provides participants in doubt with a means to avoid appearing prejudiced by assigning uncertain target items to the positive side of the screen. To shed more light on and further validate the G parameter, we manipulated the base rates of the pleasant and unpleasant items in Study 3. Presumably, if G reflects a true response bias, then it should bias toward the response that appears most frequently.

Study 3

The results of Studies 1 and 2 showed that AC and controlled processes (D and OB) varied meaningfully across dimensions on which they can be expected to vary, but they did not vary when they should not. The main goal of Study 3 was to test the validity of the guessing parameter, G. For this purpose, we attempted to manipulate this parameter by varying the number of right-hand or left-hand responses required in the task. Specifically, we manipulated the ratio of pleasant to unpleasant words appearing in a standard flowers–insects IAT. In one condition, participants saw three times as many unpleasant as pleasant words. In the other condition, this ratio was reversed. Because unpleasant and pleasant words are always assigned to the same sides of the screen, the skewed base rates should produce differences in the response tendencies assessed by the guessing parameter, G. More precisely, we expected participants to exhibit a stronger right-hand bias when they have to respond more often with the right-hand key (i.e., more pleasant words) than when they have to respond more often with the left-hand key (i.e., more unpleasant words).

Method

Thirty-seven undergraduates participated in exchange for partial course credit. All participants completed a standard flowers–insects IAT, with pleasant and unpleasant words and names of insects and flowers as stimuli. The IAT included a 20-trial unpleasant–pleasant practice block, a 20-trial insect–flower practice block, a 40-trial compatible block, a 20-trial flower–insect practice block, and a 40-trial incompatible block. Seventeen participants saw three times as many pleasant words as unpleasant words in both the practice and the test blocks (i.e., more right-hand items). Twenty participants saw three times as many unpleasant as pleasant words (i.e., more left-hand items).

Results

The overall error rate in the IAT was 5%. Because we did not expect the present base-rate manipulation to affect any parameters other than G, all other parameters (AC, D, OB) were set equal across the two conditions. The quad model fit the data with these constraints in place, $\chi^2(9) = 12.21$, $p = .20$. The parameter estimates for the quad model are presented in Table 3. As predicted, the G parameter for those participants who saw more right-hand items was significantly higher than the G parameter for those participants who saw more left-hand items, $\chi^2(1) = 10.29$, $p = .001$. For the participants who saw more right-hand items, the G parameter was significantly higher than .5, $\chi^2(1) = 11.02$, $p < .001$, indicating a significant right-hand bias. However, the G parameter for participants who saw more left-hand items was not significantly lower than .5, $\chi^2(1) = 0.84$, $p = .36$.

Unlike the AC parameters in Study 1, the AC parameters for this flowers–insects IAT did not differ from each other, $\chi^2(1) = 0.01$, $p = .92$. However, consistent with the assumption that OB may be more important for target than for attribute items, the two OB parameters were significantly different. The likelihood that bias was overcome on target items was significantly higher than the likelihood that it was overcome on attribute items, $\chi^2(1) = 10.40$, $p < .001$.

Discussion

The results of Study 3 provide further evidence for the stochastic and construct validity of the quad model, specifically for the G parameter. A manipulation of the base rates of different types of responses resulted in the predicted differences in G, the guessing parameter. These results suggest that G represents a response bias that can be influenced by features of the response environment. In the present study, participants showed a stronger right-hand bias when they had to respond more often with their right hand than when they had to respond more often with their left hand. Whereas the induced right-hand bias in the former condition was statistically significant, the left-hand bias in the latter condition was not statistically significant. The latter finding may be because of a natural right-hand bias, which is only diluted by the present left-hand manipulation, rather than reversed.

Study 4

The main goal of Study 4 was to investigate how the different processes proposed by the quad model are reflected in the scores resulting from standard data analytic strategies. For this purpose, we examined the relationship between individual parameters and standard IAT scores. Specifically, we used individual participants' estimates of AC, OB, D, and G obtained in a Black–White IAT to predict conventional latency difference scores in the same IAT. Drawing on the proposed interpretation of the parameters, we predicted that association activation (AC) should be positively related to standard IAT scores. That is, the stronger the activation of automatic associations, the greater should be the difference in response latencies between the compatible and the incompatible block. In contrast, the likelihood with which participants successfully overcome their associations (OB) should be negatively related to standard IAT scores. That is, the more likely participants are to overcome their biases, the smaller should be the difference in response latencies between the compatible and the incompatible block. With regard to D and G, we did not have any particular predictions. However, these parameters were nevertheless included in the present analyses to explore how standard IAT scores may be related to the discriminability of the stimuli or systematic guessing biases.

Method

Forty-two undergraduates participated in exchange for partial course credit. All participants completed a Black–White IAT identical to the standard IAT used in Study 2a.

Results

Parameter estimates. The quad model fit the data, $\chi^2(2) = 3.26$, $p = .20$, with an overall IAT error rate of 6%. Parameter estimates for Study 4 are presented in Table 4. G did not differ from .5, $\chi^2(1) = 1.04$, $p = .31$. More important, the AC parameter reflecting the Black–unpleasant association differed significantly from zero, $\chi^2(1) = 17.68$, $p < .001$, as did the White–pleasant association, $\chi^2(1) = 58.16$, $p < .001$. Replicating the pattern of Study 2, AC for the White–pleasant association was significantly higher than AC for the Black–unpleasant association, $\chi^2(1) = 5.30$, $p = .02$. Again, consistent with the notion that OB may be more important for target than for attribute items, the OB parameter was higher for Black and White names than for positive and negative words, $\chi^2(1) = 5.34$, $p = .02$. Whereas OB for attribute items did not differ from zero, $\chi^2(1) = 0.00$, $p = 1.00$, OB for names was significantly higher than zero, $\chi^2(1) = 5.34$, $p = .02$.

IAT latency scores. Following the "old" scoring algorithm proposed by Greenwald et al. (1998), we excluded the first two

Table 3
Parameter Estimates for the Flowers–Insects IAT, Study 3

Parameter	Comparison	Estimate	
AC	Insect–unpleasant	0.11	
	Flower–pleasant	0.11	
OB	Attributes	0.00	
	Names	0.53	
D		0.95	
		More right-hand items	More left-hand items
G		0.81	0.42

Note. Goodness of model fit: $\chi^2(9) = 12.21$, $p = .20$. IAT = Implicit Association Test; AC = association activation; OB = overcoming bias; D = discriminability; G = guessing.

Table 4
Parameter Estimates for the Black–White IAT, Study 4

Parameter	Comparison	Estimate
AC	Black–unpleasant	0.04
	White–pleasant	0.09
OB	Attributes	0.00
	Names	0.52
D		0.93
G		0.54

Note. Goodness of model fit: $\chi^2(2) = 3.26$, $p = .20$. IAT = Implicit Association Test; AC = association activation; OB = overcoming bias; D = discriminability; G = guessing.

trials of each of the compatible and incompatible blocks. Latencies higher than 3,000 ms (0.7%) were replaced with 3,000, and latencies lower than 300 ms (0.1%) were replaced with 300. Error trials were excluded from the calculation of standard IAT scores. IAT latency scores were calculated by subtracting the mean response latency on the compatible block from the mean response latency on the incompatible block, with higher values indicating a stronger preference for Whites over Blacks ($M = 132.59$, $SD = 152.09$). In order to investigate potential differences in the relation of the parameters to different scoring procedures (e.g., Mierke & Klauer, 2003), we additionally calculated IAT scores according to the new scoring algorithm presented by Greenwald, Nosek, and Banaji (2003). As with the old algorithm, higher values indicate a stronger preference for Whites over Blacks ($M = 0.45$, $SD = 0.43$). The two IAT scores were highly correlated ($r = .82$, $p < .001$).

Individual estimates. In order to investigate how the processes proposed by the quad model are reflected in standard IAT scores, we calculated parameter values for each participant on the basis of errors on the IAT. The means and standard deviations of these individual parameter estimates are displayed in Table 5.[5] Individual estimates for the six parameters (i.e., AC Black–unpleasant, AC White–pleasant, OB for attributes, OB for names, D, G) were simultaneously regressed on IAT scores calculated according to the old algorithm as well as on IAT scores calculated according to the new algorithm (see Table 6). Consistent with our predictions, both old and new IAT scores showed a positive relationship to the two association activation parameters (AC). Of most interest, both association parameters contributed independently to standard IAT scores. This result indicates that standard IAT scores reflect two

Table 5
Means and Standard Deviations for Individual Parameter Estimates for the Black–White IAT, Study 4

		Estimate	
Parameter	Comparison	*M*	*SD*
AC	Black–unpleasant	0.21	0.30
	White–pleasant	0.38	0.39
OB	Attributes	0.59	0.38
	Names	0.75	0.35
D		0.92	0.05
G		0.51	0.33

Note. IAT = Implicit Association Test; AC = association activation; OB = overcoming bias; D = discriminability; G = guessing.

Table 6
Multiple Regression IAT Scores, Study 4

		IAT old algorithm		IAT new algorithm	
Term	Comparison	β	*p*	β	*p*
AC	Black–unpleasant	.37	.023	.42	.008
	White–pleasant	.40	.031	.48	.008
OB	Attributes	−.16	.357	−.44	.014
	Names	−.38	.022	−.34	.031
D		.27	.073	.24	.097
G		.07	.629	.13	.344

Note. R^2 adjusted = .18 for old algorithm, R^2 adjusted = .24 for new algorithm. IAT = Implicit Association Test; AC = association activation; OB = overcoming bias; D = discriminability; G = guessing.

independent associations. In the present case, these are in-group favoritism and out-group derogation. OB also showed the predicted negative relationships to both the old and the new algorithm. Higher likelihood of OB on names was significantly related to lower IAT scores. A similar pattern was observed for the OB parameter for attribute items, but it failed to reach significance for the old algorithm. Interestingly, D showed a marginally significant positive relation to both old and new IAT scores. This result suggests that accurate identification of the stimuli may be associated with an increase of standard IAT scores. G was not significantly related to IAT scores.

Discussion

Results from Study 4 further support the validity of the parameters of the quad model. In addition, the present findings indicate that the parameters of the quad model offer useful insights into how the different processes proposed by the quad model are reflected in standard IAT latency scores. Consistent with our predictions, individual estimates of AC and OB were significantly related to standard IAT latency scores. AC, the association activation parameter, exhibited a positive relationship with standard IAT scores, and this relation held for both the activation of White–pleasant associations and for the activation of Black–unpleasant associations. OB, the likelihood of overcoming bias, was negatively related to standard IAT latency scores. The more participants were successful in overcoming their automatic bias, the lower were their standard IAT scores. Drawing on these findings, we suggest that one value of the parameters lies in their ability to describe the contribution of different processes more specifically and how these processes contribute to the overall performance in a given task.

Study 5

Studies 1–4 provide first evidence for the stochastic and construct validity of the parameter estimates of the quad model. However, all of these studies concern the model's relationship to

[5] Note that whereas the mean values in Table 5 are based on individual estimates, the overall estimates in Table 4 are based on aggregated error rates. Hence, the *absolute* size of the parameter values may differ; the *relative* size of the parameter values, however, should be equal.

the IAT. In Study 5, we turned from the IAT to a sequential priming measure to demonstrate the breadth of the model's applicability and its ability to shed new light on existing empirical findings. In a sequential priming task, a prime stimulus is presented for a brief period, followed by a target stimulus that has to be categorized (cf. Neely, 1977). The extent to which the prime facilitates categorization of the target is a measure of the extent to which the two concepts are linked. Study 5 applied the quad model to a particular variant of a sequential priming task designed to assess automatic associations between Blacks and guns (Payne, 2001). In this task, participants were primed with either a Black or a White face and then asked to indicate whether a subsequently presented object is a gun or a tool. The general finding in this paradigm is that participants are faster at responding to guns when they are primed with Black rather than with White faces and that they are faster at responding to tools when they are primed with White rather than Black faces (e.g., Amodio et al., 2004; Judd, Blair, & Capleau, 2004; Lambert et al., 2003; Payne, 2001; Payne, Lambert, & Jacoby, 2002). The main goal of Study 5 was to test the quad model's applicability to sequential priming paradigms, such as Payne's (2001) weapon identification task, and to demonstrate the stochastic independence of the D and OB parameters.

Other Models of Process Dissociation

A second goal of Study 5 was to compare the quad model with less complex models of process dissociation. Payne's (2001) weapon identification task is often analyzed with Jacoby's models of process-dissociation (Jacoby, 1991; Lindsay & Jacoby, 1994). However, we argue that both the C-first and the A-first models are limited in their capability to describe the full range of processes that are relevant in this task.

The C-first model (Jacoby, 1991) covers the successful identification of an object as a gun or a tool in its C parameter and response biases in guessing the nature of an object in its A parameter. More important, the A parameter reflects a particular kind of bias that drives responses only if controlled identification fails. Thus, the C-first model does not capture the influence of automatically activated associations (e.g., between Blacks and guns) that may interfere with a controlled identification of the object.

Such automatic associations are captured by the A-first model (Lindsay & Jacoby, 1994). This model covers the successful identification of an object in its C parameter and the influence of automatically activated associations on overt responses in its A parameter. However, as outlined in the introduction, the A parameter in the A-first model reflects the joint likelihood that an automatic association is activated (AC) and that the automatically activated association is not overcome (1 – OB). The likelihood that an automatic association is activated (AC) but successfully overcome (OB) is not distinguished from the case in which an automatic association is not activated in the first place (1 – AC). Hence, the model cannot distinguish between individuals who genuinely do not associate Blacks with guns and individuals who strongly associate the two but who are also successful in overcoming these associations. Moreover, the A-first model does not include a parameter for guessing the correct response if controlled identification failed. As such, the model is unable to describe general response tendencies that have been shown to play an important role in the task (e.g., Payne, 2001).

We argue that an adequate description of responses in the weapon identification task requires a consideration of all four processes postulated in the quad model. A police officer's split-second decision of whether or not to pull the trigger in response to a Black man's holding an ambiguous object can be influenced by an automatic association between Black men and guns (AC), the discriminability of the object (D), the officer's ability to overcome his or her automatic associations (OB), and, when all else has failed, a tendency to assume that he or she is threatened (G). As such, we predict that the quad model will provide a more nuanced description of the data obtained in the weapon identification task than less complex models of process dissociation.

The Role of Private Versus Anticipated Public Contexts

A third goal of Study 5 was to demonstrate the quad model's ability to shed new light on existing empirical findings. For this purpose, we reanalyzed data on the impact of accountability manipulations on responses in the weapon identification task (Lambert et al., 2003). The effects of accountability on stereotyping and prejudice have interested researchers for a variety of reasons (for a review, see Lerner & Tetlock, 1999). The most obvious of these, perhaps, is the perception of accountability as a "social panacea." Making prejudiced people accountable for their socially undesirable views should cause them to behave in a less prejudiced manner. Ironically, however, research has shown that when participants anticipate discussing their responses on a prejudice-related task, they actually exhibit more prejudice than when they think their responses are confidential (Lambert, Cronen, Chasteen, & Lickel, 1996).

Lambert et al. (2003) suggested two possible explanations for this effect. The first is a habit-strengthening or drive-based explanation. According to this explanation, accountability (or the anticipation of accountability) increases arousal, and this arousal leads to an increase in the dominant response (Hull, 1943; Zajonc, 1965). Hence, public contexts may lead to a higher activation level of habitual automatic associations than would private contexts. The second possible explanation for the increase in prejudiced responses under public conditions is an impairment-of-control account. According to this hypothesis, the anticipation of accountability decreases cognitive resources, which, in turn, decreases the ability to engage in controlled processing to combat or conceal prejudiced responses.

In order to test these alternate accounts, Lambert et al. (2003) used Payne's (2001) weapon identification task. Applying Jacoby's (1991) C-first model to this task, Lambert et al. found a decrease in the C parameter under public as compared with private conditions. However, there was no effect of the public–private manipulation on the A parameter. Drawing on these findings, Lambert et al. concluded that the anticipated public condition resulted in impairment of control but not in habit strengthening.

As outlined above, however, the C parameter in the C-first model reflects successful identification of an object, whereas the A parameter reflects biases in guessing an object given that controlled identification fails. As such, the C-first model is suitable to test the impairment-of-control account, which predicts a decrease in the controlled identification of an object. However, it seems less suitable to test the habit-strengthening account that predicts an increase in automatic associations that may interfere with controlled responding. The latter would require an application of the

A-first model (Lindsay & Jacoby, 1994), which covers the successful identification of an object in its C parameter and the influence of automatically activated associations on overt responses in its A parameter. However, because the A parameter in the A-first model reflects the joint likelihood that automatic bias is activated (AC) and automatic bias does not drive the response (1 – OB), the model cannot distinguish between people who genuinely do not associate Blacks with guns and individuals who strongly associate the two but who are also successful in overcoming these associations.

This seems particularly relevant in the present case of comparing public and private contexts. Specifically, one could argue that automatic associations between Blacks and guns were indeed activated to a greater degree in Lambert et al.'s (2003) public conditions. However, these associations may also be overcome more often, given that participants may have a higher motivation to overcome their biases in public contexts. This differential influence on AC and OB cannot be identified with the A-first model. Moreover, the A-first model does not include a parameter for guessing the correct response if controlled identification fails, which may also play a significant role in weapon identification (cf. Payne, 2001). Hence, even the A-first model seems unable to describe the full range of processes contributing to participants' performance in the weapon identification task and how these processes are affected by public versus private contexts. In Study 5, we reanalyzed the data from Lambert et al. (2003) with the quad model in order to better understand exactly how the different kinds of automatic and controlled processes were affected by the public versus private manipulation.[6]

Method

One hundred twenty-seven undergraduates participated in the experiment. The priming task has been described in detail by Payne (2001), so we simply summarize it here. The primes for the task included photographs of four White faces and four Black faces, half of which were male and half of which were female. The target stimuli were photographs of four handguns and four hand tools. Each trial consisted of a 500-ms presentation of a pattern mask, followed by a 200-ms presentation of a prime face, and an immediate presentation of a target stimulus. The task was to indicate, through a button press, whether the target stimulus was a gun or a tool. The target was presented for 100 ms and was followed by a 450-ms presentation of a visual mask. If participants failed to respond before the end of the mask presentation, then they were warned that they had not responded quickly enough. Participants completed 48 practice trials, after which they were told either that their responses were confidential (private condition) or that they were expected to discuss their responses with the other participants in the testing room (public condition). After the manipulation, participants completed an additional three blocks of the priming task, each 128 trials long, for a total of 384 observations per participant.

Results

A description of the structure of the quad model as well as the C-first and the A-first models for these data are provided in Appendix C. The data had an overall error rate of 21%, substantially higher than the error rates for the IATs. The quad model fit the data relatively well, $\chi^2(3) = 5.17$, $p = .16$, slightly better than the C-first model did, $\chi^2(6) = 6.71$, $p = .35$. The A-first model, $\chi^2(6) = 13.25$, $p = .02$, did not show a satisfactory fit to the data.

Parameter estimates are provided in Table 7. Consistent with Lambert et al.'s (2003) analyses using Jacoby's (1991) C-first model, participants in the private condition were more capable of determining the correct response than were those in the public condition. This effect is reflected in a significant difference in the D parameter for public versus private conditions, $\chi^2(1) = 51.47$, $p < .001$. Also consistent with Lambert et al.'s analyses, there were no differences in G as a function of public versus private conditions. In the present analyses, this lack of an effect is indicated by an adequate model fit when G was set equal across the experimental conditions. Still, the G parameter was significantly lower than .50, $\chi^2(1) = 23.22$, $p < .001$, indicating a general bias toward guessing "gun." This bias possibly reflects a tendency toward assuming that there is danger when the situation is ambiguous.

Table 7
Parameter Estimates for the Sequential Priming Task Used by Lambert et al. (2003), Study 5

Parameter	Prime	Estimate: Public context	Estimate: Private context
AC	White-male	0.11	0.01
	White-female	0.05	0.00
	Black-male	0.09	0.03
	Black-female	0.00	0.01
OB		1.00	0.00
D		0.55	0.63
G		0.46	

Note. Goodness of model fit: $\chi^2(3) = 5.17$, $p = .16$. AC = association activation; OB = overcoming bias; D = discriminability; G = guessing.

In this study, AC reflects an automatic association between the racial category and the category-congruent item (i.e., Black–gun and White–tool). An increase in AC in the anticipated public as compared with the private condition would provide evidence for habit strengthening. In fact, there was a general increase of AC as a function of the context, $\chi^2(4) = 11.20$, $p = .02$. More careful analysis revealed that this difference was driven primarily by gun–tool associations with male targets, which were significantly higher in the public condition, $\chi^2(2) = 8.24$, $p = .01$. Associations with female targets did not differ across conditions, $\chi^2(2) = 3.81$, $p = .14$; and did not differ from zero in either the anticipated public, $\chi^2(2) = 3.29$, $p = .19$, or the private condition, $\chi^2(2) = 0.49$, $p = .78$. This finding is understandable in light of the fact that ethnic stereotypes are largely based on male members of the groups (Eagly & Kite, 1987).

Finally, OB was higher in the public than in the private condition, although this difference failed to reach significance, $\chi^2(1) = 1.77$, $p = .18$. However, with low estimates of AC, such as those observed in the present study, the power to test hypotheses about OB is fairly low. Nevertheless, the obvious increase (1 vs. 0) in OB is consistent with the assumption that participants may be more motivated to overcome their automatic biases in public as compared with private contexts.

Discussion

Results from Study 5 indicate that the quad model can be applied not only to the IAT but also to other tasks involving an

[6] We thank Alan Lambert for generously sharing his data and expertise.

interplay between the four distinct processes proposed by the model. Moreover, the present findings indicate that a consideration of all four processes postulated by the quad model provides a better description of the responses in the weapon identification task (Payne, 2001) than less complex models of process dissociation (Jacoby, 1991; Lindsay & Jacoby, 1994). The present findings demonstrate that the quad model can provide a more fine-grained description of data than can other models of process dissociation. In the present case, some of the findings obtained with the quad model mirror those of standard process-dissociation models. Specifically, it seems that D of the presented stimuli is generally lower under public than under private conditions. This conclusion is supported by a decrease in the C parameter of the C-first model, used by Lambert et al. (2003), and a corresponding decrease in the D parameter of the quad model.[7] This finding is consistent with the impairment-of-control account, suggesting that participants' ability to discriminate the stimuli decreases under anticipated public as compared with private conditions. More important, however, with the quad model, effects that were not discovered by other models were able to be teased apart. Specifically, whereas the controlled identification (D) of guns and tools decreased in public contexts, success at overcoming automatic associations (OB) showed a tendency in the opposite direction. This result suggests that, even though public contexts reduced the ability to correctly identify the presented objects, participants' motivation to overcome their automatic associations between Blacks and guns increased under public conditions. Moreover, we also found greater activation levels of automatic associations under public as compared with private conditions. This finding is consistent with a habit-strengthening account, suggesting that associations (or dominant responses) are more likely to be activated under public as compared with private conditions (Zajonc, 1965). Taken together, these results indicate that the quad model provides a clearer picture of the complexity of the differences between public and private conditions, suggesting that both impairment of control and habit strengthening affected performance in Lambert et al.'s (2003) study.

General Discussion

In the present article, we proposed and tested the quad model, which measures the influence of multiple distinct processes on implicit task performance. Specifically, we argued that performance on implicit measures is influenced by at least four different processes: the automatic activation of an association (*association activation*), the ability to determine a correct response (*discriminability*), the success at overcoming automatically activated associations (*overcoming bias*), and the influence of response biases that may influence responses in the absence of other available guides to response (*guessing*). The quad model provides a mathematical tool to disentangle the contribution of all four of these processes. This approach may allow us to learn more about the specific features of each of these processes as well as their interactions with the processing context.

Validity of the Quad Model

The quad model is a multinomial model that estimates parameters on the basis of observable error rates. Validity of a multinomial model is usually based on two kinds of evidence: stochastic independence of the parameters and meaningful variation in the parameters. To demonstrate stochastic independence, we have shown that the four parameters can vary independently of each other. To demonstrate the construct validity of the parameters, we have shown that they vary in concert with the processes we suppose they represent and that they respond appropriately to various manipulations.

Stochastic validity. In each study, AC varied as a function of the association (e.g., Black–unpleasant) but not as a function of the stimulus type (i.e., target vs. attribute). This is in contrast to OB, which varied as a function of stimulus type in Studies 3 and 4 but not as a function of association. In Study 2, AC was affected by the association but not by the manipulation of time constraints. These results suggest that AC can vary independently of D and OB, both of which were affected by the ability to use control. Moreover, in Study 3, G, but not AC, varied as a function of the number of required left-hand or right-hand responses, supporting the independence of AC and G. The D and OB parameters varied independently as a function of the public–private manipulation in Study 5. Specifically, a public context led to a decrease in D but to an increase in OB. Finally, in Study 3, G, but not OB and D, was affected by the number of required left-hand or right-hand responses, supporting the independence of those two parameters from G.

Construct validity. Studies 2–5 also were concerned with establishing the construct validity of the parameters. Specifically, we tried to show that the parameters actually assess the individual processes we think they reflect. The construct validity of the AC parameter was demonstrated by a number of findings. In Study 2, the manipulation of time constraints had no effect on AC, suggesting that it reflects an automatic process. The main effect of AC in the prediction of IAT latency scores in Study 4 also is consistent with our interpretation of the parameter as a measure of association strength. Finally, Study 5 showed that AC was enhanced in the public condition. This result is consistent with drive-based models, suggesting that public contexts increase the influence of dominant responses (e.g., Zajonc, 1965), thus supporting our view of AC as reflecting automatically activated associations.

Construct validity of the discriminability parameter (D) was supported by a variety of findings. First, in Study 2, D was diminished by time constraints, indicating that it reflects a process requiring effort and attention. Second, our reanalysis of Lambert et al.'s (2003) data in Study 5 showed that D was lower in the anticipated public than in the private condition. This finding is consistent with Lambert et al.'s interpretation of the results in terms of impairment of control, which suggests that heightened arousal in the anticipated public condition leads to diminished attention.

Construct validity of the overcoming bias parameter (OB) was also established in multiple ways. First, Study 2 showed that OB was diminished by time constraints, supporting our interpretation of this parameter as an effortful process. Second, in Study 4, OB

[7] Note that even though the D parameter of the quad model and the C parameter of the C-first model (Jacoby, 1991) are conceptually similar, they are not identical. Whereas the D parameter of the quad model reflects the likelihood that an object *can* be identified, the C parameter of the C-first model reflects the likelihood that an object *is* identified.

was negatively related to IAT latency scores. As the likelihood that participants were successful at overcoming their biases increased, the amount of implicit prejudice reflected in their standard IAT scores decreased. Finally, our reanalysis of Lambert et al.'s (2003) data in Study 5 showed that OB tended to increase when participants believed they would be publicly accountable for their behavior. Not surprisingly, when participants expected their behavior to be evaluated by others, they were more concerned about not exhibiting socially unacceptable biases, and they worked harder to ensure that outcome. Taken together, these findings support our interpretation of OB as a controlled process that is influenced by both motivation and ability.

Finally, the construct validity of G was supported in Study 3, which showed that G varied in concert with the base rates of required right-hand and left-hand responses. Participants who had to make more right-hand responses showed a greater right-hand bias in guessing than did participants who had to make more left-hand responses.

Implications and Future Directions

Emerging research in social psychology is demonstrating the complexity of the cognitive processes that produce even simple, discrete responses. A police officer's split-second decision of whether or not to pull the trigger in response to a Black man's holding an ambiguous object (e.g., Correll et al., 2002; Greenwald, Oakes, & Hoffman, 2003; Payne, 2001) may be influenced by an implicit association between Black men and guns (AC), the discriminability of the object (D), the officer's capacity to overcome an automatic bias and not shoot if the man is determined not to have a gun (OB), and, when all else has failed, a tendency to assume that he or she is threatened (G). In our eyes, conventional strategies to disentangle automatic and controlled processes by means of explicit and implicit measures are inadequate for the full understanding of such complex responses. Moreover, processes typically assumed to be similar in the amounts of cognitive resources they consume may have very different meanings, even within the context of the same response. In Study 5, for instance, D and OB, two processes that can be described as controlled, were influenced differently by the same context manipulation. The quad model goes beyond a simple quantitative differentiation between automaticity and control by allowing the exploration of differences in the actions of multiple distinct processes that may or may not be similar in the relative amounts of resources they consume. Drawing on these considerations, we argue for a new perspective that focuses on qualitatively distinct processes rather than on quantitative differences in automatic and controlled processing models.

The quad model also offers some interesting perspectives for future research. In the literature on prejudice, for example, there is a brewing debate over the interpretation of implicit measures of prejudice, with some researchers arguing that low scores reflect weak attitudes (e.g., Blair et al., 2001; Dasgupta & Greenwald, 2001; Kawakami et al., 2000; Rudman, Ashmore, & Gary, 2001), and others suggesting that low scores may reflect skill at self-regulation (e.g., Devine & Monteith, 1999; McFarland & Crouch, 2002; Moskowitz et al., 1999). These interpretations have very different implications for reducing prejudiced behavior. Would we be best served by exposing people to positive role models (e.g., Blair et al., 2001; Dasgupta & Greenwald, 2001) or by training people to overcome their biases (e.g., Kawakami et al., 2000; Moskowitz et al., 1999)? From the perspective of the quad model, it seems possible that both approaches may have important effects on prejudiced responses. However, our findings suggest that researchers should exercise caution in assuming that the implicit prejudice scores they calculate with priming measures or with the IAT reflect exclusively the strength of automatic associations. Clearly, attempts to overcome these associations also contribute to performance on these tasks. What the quad model offers is a means to independently assess these processes and gauge their joint contributions to behavior.

Another interesting question for future research concerns the weak relationship between different kinds of implicit measures. Cunningham et al. (2001), for example, found that the correlation between the IAT (Greenwald et al., 1998) and affective priming (Fazio et al., 1995) is only moderate, even when measurement error is controlled. This finding may indicate that the nature of the associations assessed with these tasks have only partial overlap (e.g., Olson & Fazio, 2003). From the perspective of the quad model, however, one could argue that the automatic associations assessed with these measures may be the same but that the tasks are differentially affected by other processes such as, for example, OB.

The quad model's ability to measure association strength and self-regulation simultaneously is also relevant in other areas of psychology. With regard to clinical applications, for example, Teachman and Woody (2003) recently found evidence for treatment-related changes in patients with arachnophobia using an IAT. Specifically, these researchers found that, over the course of treatment, participants exhibited lower scores in an IAT designed to assess implicit negative evaluations of spiders. Drawing on the present findings, one could argue that such changes may reflect either a reduction in automatic negative associations or increased success in controlling automatic negativity. The latter assumption is consistent with a recent claim by De Jong, Van den Hout, Rietbroek, and Huijding (2003), who argued that nonphobic individuals may be characterized by their ability to overcome their automatic negative associations rather than by a lower level of automatic attitude activation. Future research may help to further clarify the particular role of AC and OB in phobias.

Conclusion

As a mathematical model of implicit task performance, the quad model has the potential to extend our understanding of the interplay of multiple processes in the measurement of automatic associations. The present application to two different types of measures, the IAT and a sequential priming task, shows that these measures are far from process pure. Responses on these measures reflect not only the strength of the association but also participants' ability to discriminate the stimuli, their ability to overcome their automatic associations, and general guessing biases. However, we suggest that the lack of process purity in these measures is an asset rather than a flaw. Using the quad model, we can dissociate components of tasks that are more similar to those we may encounter in the real world and observe their behavior across settings and individuals. Accordingly, we propose a conceptualization of processing that includes not just two quantitatively different processing modes but four qualitatively distinct processes. In our view, methods that take into account multiple processes' influences on overt responses are a desirable alternative to methods that attempt to eliminate the influence of all but a single process.

References

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 & T. K. Srull (Eds.), *Handbook of social cognition. Vol. 1: Basic processes* (2nd ed., pp. 1–40). Hillsdale, NJ: Erlbaum.

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

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

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues, 55,* 429–444.

Buchner, A., Erdfelder, E., & Vaterrodt-Pluennecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General, 124,* 137–160.

Chaiken, S., & Trope, Y. (1999). (Eds.). *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.

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

Cunningham, W. A., 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). A structural analysis of indirect measures of attitudes. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 219–244). Mahwah, NJ: Erlbaum.

De Jong, P. J., Van den Hout, M. A., Rietbroek, H., & Huijding, J. (2003). Dissociations between implicit and explicit attitudes toward phobic stimuli. *Cognition & Emotion, 17,* 521–545.

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

Devine, P. G., & Monteith, M. J. (1999). Automaticity and control in stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 339–360). New York: Guilford Press.

Eagly, A. H., & Kite, M. E. (1987). Are stereotypes of nationalities applied to both women and men? *Journal of Personality and Social Psychology, 53,* 451–462.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). San Diego, CA: Academic Press.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69,* 1013–1027.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50,* 229–238.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). San Diego, CA: Academic Press.

Gawronski, B., & Bodenhausen, G. V. (2004). *Accessibility effects on implicit social cognition: The role of accessible content and accessibility experiences*. Manuscript submitted for publication.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford Press.

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., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197–216.

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

Haddock, G., Zanna, M. P., & Esses, V. M. (1993). Assessing the structure of prejudicial attitudes: The case of attitudes toward homosexuals. *Journal of Personality and Social Psychology, 65,* 1105–1118.

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. New York: Appleton-Century-Crofts.

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

Jacoby, L. L., McElree, B., & Trainham, T. N. (1999). Automatic influences as accessibility bias in memory and Stroop tasks: Toward a formal model. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 461–486). Cambridge, MA: MIT Press.

Judd, C. M., Blair, I. V., & Chapleau, K. M. (2004). Automatic stereotypes vs. automatic prejudice: Sorting out the possibilities in the Payne (2001) weapon paradigm. *Journal of Experimental Social Psychology, 40,* 75–81.

Kawakami, K., Dion, K. L., & Dovidio, J. F. (1999). The Stroop task and preconscious activation of racial stereotypes. *Swiss Journal of Psychology, 58,* 241–250.

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.

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the "Who said what?" paradigm. *Journal of Personality and Social Psychology, 75,* 1155–1178.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus–response compatibility: A model and taxonomy. *Psychological Review, 97,* 253–270.

Lambert, A. J., Cronen, S., Chasteen, A. L., & Lickel, B. (1996). Private vs. public expressions of racial prejudice. *Journal of Experimental Social Psychology, 32,* 437–459.

Lambert, A. J., Payne, B. K., Jacoby, L. L., Shaffer, L. M., Chasteen, A. L.,

& Khan, S. R. (2003). Stereotypes as dominant responses: On the “social facilitation” of prejudice in anticipated public contexts. *Journal of Personality and Social Psychology, 84,* 277–295.
Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125,* 255–275.
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.
McConahay, J. B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. D. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). San Diego, CA: Academic Press.
McFarland, S. G., & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition, 20,* 483–510.
Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology, 85,* 1180–1192.
Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77,* 167–184.
Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.
Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19,* 625–666.
Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science, 14,* 636–639.
Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81,* 181–192.
Payne, B. K., Lambert, A. J., & Jacoby, L. L. (2002). Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. *Journal of Experimental Social Psychology, 38,* 384–396.
Petty, R. E., & 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.
Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95,* 318–339.
Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist, 9,* 1043–1056.
Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology: General, 133,* 139–165.
Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). “Unlearning” automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology, 81,* 856–868.
Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84,* 1–66.
Sherman, J. W., Groom, C. J., Ehrenberg, K., & Klauer, K. C. (2003). Bearing false witness under pressure: Implicit and explicit components of stereotype-driven memory distortions. *Social Cognition, 21,* 213–246.
Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology, 75,* 589–606.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84,* 127–190.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22.
Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality & Social Psychology Review, 4,* 108–131.
Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of human behavior. *Personality and Social Psychology Review, 8,* 220–247.
Stroop, J. R. (1935). Studies on the interference in serial verbal reactions. *Journal of Experimental Psychology, 59,* 239–245.
Teachman, B. A., & Woody, S. R. (2003). Automatic processing in spider phobia: Implicit fear associations over the course of treatment. *Journal of Abnormal Psychology, 112,* 100–109.
Toth, J. P., Reingold, E. M., & Jacoby, L. L. (1994). Toward a redefinition of implicit memory: Process dissociations following elaborative processing and self-generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 290–303.
Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review, 93,* 239–257.
Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review, 101,* 34–52.
Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D. T. Gilbert, S. E. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 446–496). New York: McGraw-Hill.
Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72,* 262–274.
Zajonc, R. B. (1965, July 16). Social facilitation. *Science, 149,* 269–274.
Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35,* 151–175.

(*Appendixes follow*)

Appendix A

Fitting the Quad Model to Data

The processes measured by the quad model are relevant to any implicit measure that is based on the logic of response compatibility (cf. De Houwer, 2003; Kornblum et al., 1990). However, the quad model can be applied only to carefully designed tasks that have more uniquely predicted categories of responses than estimated parameters. In the IATs used in this article, error rates on the compatible and incompatible blocks for each of the four item types ("White" names, "Black" names, pleasant words, unpleasant words) add up to eight categories, but the equations used to predict White and pleasant items and Black and unpleasant items are the same, so the compatible block provides only two unique categories. Error rates from the practice blocks provide two additional categories, so there is a total of eight response categories. In these IATs, we estimated six parameters: two AC parameters, one D parameter, two OB parameters, and one G parameter. The difference between the number of response categories and the number of parameters is the number of degrees of freedom for the model—in this case 2. More details about the specific structures of the models used for the IAT are provided in Appendix B.

The model can also be fit to sequential priming tasks. In the priming task used in Study 5, there were eight observable categories of responses. The four prime types, Black and White female and male faces, were paired with gun and tool targets for each participant, adding up to eight response categories. When the data were separated by the public and private conditions, there were 16 categories. With six parameters for each condition (four AC parameters, one OB parameter, one D parameter, plus a single G parameter that collapses across the public and private conditions), the model has three degrees of freedom. More details about the specific structures of the models used in Study 5 are provided in Appendix C.

Another important aspect of model fit is the unique identifiability of the model. *Identifiability* refers to the mapping of a unique set of parameters onto each unique set of observed responses. The quad model proved to be identifiable in each of the contexts presented here. In the present studies, identifiability was proven by mathematically inverting the quad model for Study 1. Because the particular model structures used in all of the remaining studies represent constrained versions of the model in Study 1, all of these models can be considered identifiable. As an additional test of identifiability, we fit the same data to the model several times and, in each case, obtained the same parameter estimates.

Categories of Data

A category of data (e.g., the error rate for pleasant words on an incompatible block) is considered "uniquely" predicted if the equation associated with it in the quad model is unique. For instance, the model predicts that a pleasant word will be assigned to the correct side of the screen in the incompatible block with the probability

$$p(\text{correct}|\text{pleasant, incompatible}) = AC \times D \times OB + (1 - AC) \times D + (1 - AC) \times (1 - D) \times G.$$

This equation sums the three possible paths by which a correct answer can be returned in this case. The first part of the equation, $AC \times D \times OB$, is the likelihood that the association is activated and that the correct answer can be discriminated and that the association is overcome in favor of controlled responding. If the association is overcome, the correct response is returned. 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 determined. In this case, the correct response is returned. Finally, $(1 - AC) \times (1 - D) \times G$ is the likelihood that the association is not activated, the correct answer cannot be discriminated, and the participant guesses the right-hand response. For a pleasant word, this guess would return the correct response. The sum of these probabilities is the total probability of a correct response for the item.

The probability of an incorrect response is

$$p(\text{incorrect}|\text{pleasant, incompatible}) = 1 - p(\text{correct}) = AC \times D \times (1 - OB) + AC \times (1 - D) + (1 - AC) \times (1 - D) \times (1 - G).$$

Here, if the association is activated and the correct answer can be discriminated and the association is not overcome, the association will drive the response, and the incorrect response will be returned. This instance is represented by $AC \times D \times (1 - OB)$. If the association is activated and the correct response cannot be determined, the association will drive the response, and the incorrect response will be returned. This instance is represented by $AC \times (1 - D)$. Finally, if the association is not activated and the correct response cannot be determined and the participant guesses the left-hand response, then the incorrect response will be returned. This instance is represented by $(1 - D) \times (1 - G)$. Obviously, this probability of an incorrect response is completely redundant with the probability of returning a correct response. It does not add anything to our ability to describe the data, so correct and incorrect responses on a single item type in a single block type are not considered unique.

However, the probability of a correct response on a Black name on the pleasant block is considered different from the probability of a correct response on a pleasant word. Although the general equation is the same,

$$p(\text{correct}|\text{Black, incompatible}) = AC \times D \times OB + (1 - AC) \times D + (1 - AC) \times (1 - D) \times G,$$

the more correct equations in this instance of the model are

$$p(\text{correct}|\text{pleasant, incompatible}) = AC_{\text{White/pleasant}} \times D_{\text{pleasant/unpleasant}} \times OB_{\text{pleasant/unpleasant}} + (1 - AC_{\text{White/pleasant}}) \times D_{\text{pleasant/unpleasant}} + (1 - AC_{\text{White/pleasant}}) \times (1 - D_{\text{pleasant/unpleasant}}) \times G_{\text{test}}$$

and

$$p(\text{correct}|\text{Black, incompatible}) = AC_{\text{Black/unpleasant}} \times D_{\text{Black/White}} \times OB_{\text{pleasant/unpleasant}} + (1 - AC_{\text{Black/unpleasant}}) \times D_{\text{Black/White}} + (1 - AC_{\text{Black/unpleasant}}) \times (1 - D_{\text{Black/White}}) \times G_{\text{test}}.$$

The specific parameters used to estimate each of these probabilities are different. Thus, the two categories of data are considered unique. This observation is related to a broader aspect regarding the nature of modeling in general. If we were to use completely different parameters to estimate each observed value, the model would fit perfectly. The challenge in modeling is to achieve parsimony in the number of parameters and accuracy in predicting the data. Here, parsimony is enforced by the degrees of freedom: There must be fewer parameters than categories of data. Accuracy is enforced in the model fit: If the model is inadequate, the predicted error rates will differ substantially from the observed error rates.

Estimating Parameter Values

To estimate the parameter values, we create an equation for each unique category of data, with the parameter values set at arbitrary levels, usually .5. Each of these equations produces a predicted probability for the category. A chi-squared value is computed using the observed error rate and the expected error rate provided by the model. These chi-squared values are summed across all categories to produce an overall chi-squared value for the model. The smaller this value, the better, so the parameters are changed through maximum likelihood estimation (MLE) until they return a minimum possible value of the chi-squared. Because chi-squared is an inferential statistic, to say that a model "fits" the data is to accept the null. Thus, it is relatively easier to get a significant fit with fewer observations. This issue means that not only the fit but also the stochastic and construct validity of the parameters is important in evaluating and using the model.

Hypothesis Testing

To test a hypothesis about two or more parameters, we first fit the model with the parameters varying independently and then again with them set equal. The difference in the chi-squared value for the two fits is the chi-squared for the hypothesis test. The number of degrees of freedom for the test is the reduction in parameters estimated by the model. For instance, if two AC parameters are set equal, one parameter is saved, so the hypothesis test has one degree of freedom. If two OB parameters are set equal to zero simultaneously, two parameters are saved, so the test has two degrees of freedom. Unlike model fitting, power for these hypothesis tests is lower with fewer observations. We choose to focus more on what the parameters can tell us than on the overall model fits because there are fewer concerns about sample size, power, and accepting the null.

Using the Quad Model

The IAT has been our task of major choice in this article because it is sufficiently complex to resolve the conflicts between parsimony, accuracy, categories of data, and degrees of freedom. For the novice user, the IAT is a good task in which to apply the quad model because it is highly adaptable and easy to administer. An Excel spreadsheet template for applications of the quad model to IAT data has been prepared and is available on the internet at http://mypage.iu.edu/~rconrey/quad-model.html.

Appendix B

Model Specifications for Studies 1–4

The predicted responses for the quad model are described in Figure 1. The parameter estimates for the quad model are detailed in Studies 1-4, so they are only summarized here. All models of the IAT presented in this article included all blocks of the IAT, including the two-category practice blocks to estimate the parameter values. The left- and right-hand responses in the single-category practice blocks provided two additional uniquely predicted response categories.

Two AC parameters were estimated for each IAT. These represented the compatible target–attribute associations (e.g., White–pleasant and Black–unpleasant). In Study 2, these parameters were set equal across the response window and no window conditions. In each study, a single D parameter was estimated to assess the discriminability of all items. That is, the D parameters were set equal across AC and 1 – AC. Two OB parameters were estimated for each IAT, one for the attribute items and one for the target items. Separating OB for these item types allowed us to make predictions about the selective devotion of resources to one type of item over another. A single G parameter was estimated in Studies 1 and 4. In study 2, a G parameter for the practice blocks was estimated separately and set equal across the two groups to anchor the two conditions to each other. This was done to provide an additional link between the two processing trees, so that we could test for differences in G on the test blocks as well as AC between the two conditions. In Study 3, G was the only parameter allowed to vary across the two IATs.

With eight uniquely predicted categories, two AC parameters, one D parameter, two OB parameters, and one G parameter, the model for the IAT has two degrees of freedom. In Study 2, in which the AC parameters were set equal across response window conditions and a single additional G parameter for practice was estimated for both conditions, the model had five degrees of freedom.

For the sake of consistency, we maintained the same structure of the model across all four studies, but it is certainly possible to imagine cases in which it might be altered to test specific hypotheses. For instance, more than two D parameters could be estimated in a case where some items were considered more discriminable than others, or D might be allowed to vary across the AC and 1 – AC cases as discussed in Footnote 4. However, care must be taken to ensure that the model is identifiable in each of these cases.

Appendix C

Model Specifications for Study 5

Quad Model

In each condition, private and anticipated public, four AC parameters were estimated: White man–tool, Black man–gun, White woman–tool, and Black woman–gun. One D parameter was estimated for each condition: private and anticipated public. One OB parameter was estimated for each condition: private and anticipated public. Only one G parameter was estimated. G was coded to represent a bias toward guessing "tool." There were 16 uniquely predicted categories of observations. With eight AC parameters, two D parameters, two OB parameters, and one G parameter, there were three degrees of freedom for the quad model.

C-First Model

The predicted pattern of responses for the C-first model is depicted in Figure 1. Two C parameters were estimated, one for the public and one for the private condition. In addition, four A parameters were estimated in each condition: White man–tool, Black man–gun, White woman–tool, and Black woman–gun. There were 16 uniquely predicted categories of observations. With eight A parameters, two C parameters, and two unique processing trees, there were six degrees of freedom for the C-first model.

A-First Model

The predicted pattern of responses for the A-first model is depicted in Figure 2. Four A parameters were estimated in each condition: White man–tool, Black man–gun, White woman–tool, and Black woman–gun. In addition, two C parameters were estimated, one for each of the between-participants conditions. There were 16 uniquely predicted categories of observations. With eight A parameters, two C parameters, and two unique processing trees, there were five degrees of freedom for the A-first model.

Received July 8, 2004
Revision received April 28, 2005
Accepted May 10, 2005 ■