ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Experimental Social Psychology xxx (2005) xxx–xxx

Journal of Experimental Social Psychology

www.elsevier.com/locate/jesp

Decoding the implicit association test: Implications for criterion prediction ☆

Hart Blanton [a,*], James Jaccard [b], Patricia M. Gonzales [c], Charlene Christie [d]

[a] *Department of Psychology, Texas A&M University, USA*
[b] *Department of Psychology, Florida International University, USA*
[c] *New York State Energy Research and Development Authority, USA*
[d] *Indiana University–Purdue University, Columbus, USA*

Received 6 October 2004; revised 14 June 2005

Abstract

The implicit association test (IAT) is believed to measure implicit evaluations by assessing reaction times on two cognitive tasks, often termed "compatible" and "incompatible" tasks. A common rationale for studying the IAT is that it might improve our prediction and understanding of meaningful psychological criteria. To date, however, no clear psychometric theory has been advanced for this measure. We examine the theory, methods and analytic strategies surrounding the IAT in the context of criterion prediction to determine measurement and causal models a researcher embraces (knowingly or unknowingly) by using the test. Our analyses reveal that the IAT revolves around interpretation of two distinct *relative constructs*, one at the conceptual level and one at the observed level. We show that interest in relative implicit evaluations at the conceptual level imposes a causal model that is restrictive in form. We then examine measurement models of the IAT and show how computing a difference score at the observed level may lack empirical justification. These issues are highlighted in a study replicating an effect established in the literature (Study 1). We then introduce a new variant of the IAT and use it to evaluate the reasonableness of traditional IAT methods (Study 2).

Keywords: Implicit association test; Structural equation modeling; Validity; Prejudice; Discrimination

The past decade has seen a dramatic increase in the use of cognitive responses as indicators of attitudes (Bargh, Chaiken, Govender, & Pratto, 1992; De Houwer, 2003; Devine, 1989; Greenwald, McGhee, & Schwartz, 1998; Greenwald & Nosek, 2001; Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Dovidio, Kawakami, & Gaertner, 2002). A great deal of this work has focused on the use of the *Implicit Association Test* (IAT). The IAT was designed to measure individuals' underlying attitudes by examining the speed with which they can associate two different concepts with two different evaluative anchors. The primary rationale for measuring attitudes in this way, rather than using more traditional self-report measures, is that the response latencies can provide insights into a person's attitudes that self-report inventories belie. In this spirit, the IAT has been applied to a wide range of social problems, in many branches of psychology. It is thus important to examine the conceptual and psychometric underpinnings of this new measure.

We are not alone in asking questions that speak to the validity of the IAT. Numerous researchers have raised concerns regarding its methodological and theoretical limitations (Brendl, Markman, & Messner, 2001; Karpinski, 2004; Karpinski & Hilton, 2001; Klauer & Mierke, 2004; Fazio & Olson, 2003; Fiedler, Messner, & Matthias, 2004; Olson & Fazio, 2004; Rothermund & Wentura, 2001, 2004; Steffens, 2004). Our article differs from these

☆ The authors thank John T. Cacioppo, Curt Hardin, Chet Insko, Andy Karpinski, Jim Neely, Brett Pelham, Diederik Stapel, Penny Visser and the SUNY-Albany and UNC graduate students for their insightful comments and suggestions.
* Corresponding author.
E-mail address: hblanton@psych.tamu.edu (H. Blanton).

0022-1031/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2005.07.003

efforts by focusing on (1) the formal measurement models that link observed IAT scores to the theoretical IAT construct and (2) the formal causal models that link the theoretical IAT construct to the psychological criteria one hopes to predict.

Psychometrics: Linking observed variables to conceptual variables

Because psychological constructs such as attitudes cannot be observed directly, researchers who develop psychometric inventories must advance psychometric theories that explain how observable responses (e.g., responses to a questionnaire) are influenced by the unobserved latent constructs these responses are thought to measure (e.g., racial prejudice). In this regard, the IAT focuses on two distinct constructs that are *relative* in nature. At the *conceptual level*, the IAT is thought to measure the difference between two implicit evaluations (e.g., the relative implicit preference for Whites versus Blacks). At the *observed level*, an IAT score is computed by taking the difference between two behavioral responses (i.e., response latencies). We turn first to the conceptual level and show that the relative construct of the IAT invokes a causal model that is restrictive in form and that will be inappropriate in many criterion-prediction contexts. We then turn to the observed level and show how a relative construct invokes measurement models that lack empirical support.

Conceptual level: Relative causal models

The IAT is promoted as a measure that can assess the (unobservable) latent psychological construct, *implicit attitudes*. The main way in which implicit attitudes are thought to differ from explicit attitudes is a supposed tendency for the former to be inaccessible in declarative memory. A more important departure from the traditional conceptualization of attitudes has been the way in which researchers treat attitude objects. Although the definition of an "attitude" varies from theory to theory (Eagly & Chaiken, 1993; Fishbein & Ajzen, 1975; Petty, Wegener, & Fabrigar, 1997), most traditional theories conceptualize an (explicit) attitude as a bipolar dimension that reflects the degree of favorable or unfavorable evaluation of an attitude object. Attitude objects are entities that the individual can perceive or cognitively make note of, and they can refer to concrete objects, abstract concepts, and to both social and nonsocial categories. Whereas traditional attitude theories place a *single* attitude object on a bipolar evaluative dimension, IAT researchers have tended to focus their attention on *relative implicit attitudes* or *relative implicit evaluations* (Greenwald et al., 1998). Relative implicit attitudes reflect the difference between how a person implicitly evaluates two distinct attitude objects.

Greenwald and Farnham (2000) acknowledged that the IAT is designed to measure a relative construct and they suggested that "the IAT can nevertheless be effectively used because many socially significant categories form complementary pairs, such as positive–negative (valence), self–other, male–female, Jewish–Christian, young–old, weak–strong, warm–cold, liberal–conservative, aggressive–peaceful, etc." (p. 1023). This statement seems reasonable at a glance, but it ignores an unavoidable statistical limitation of relative measures. Consider as an example one of the measures that was developed in the initial paper on the IAT (Greenwald et al., 1998). This measure, which we hereafter refer to as the "race IAT," was designed to assess implicit racial attitudes by measuring the relative implicit preference for Whites over Blacks.

The most obvious way in which to conceptualize this relative construct is as the difference between how a person implicitly evaluates Whites versus Blacks.[1] This true (unmeasured and unobservable) relative implicit evaluation (RIE) can be represented as

$$\mathrm{RIE} = E_{\mathrm{W}} - E_{\mathrm{B}}, \tag{1}$$

where E_{W} is the true (unmeasured and unobservable) implicit evaluation of Whites and E_{B} is the true (unmeasured and unobservable) implicit evaluation of Blacks. Suppose that an investigator wants to predict a criterion, Y, from the relative implicit evaluation. If the researcher applies a traditional linear model, the relationship between the variables can be described as:

$$Y = \alpha + \beta\,\mathrm{RIE} + \varepsilon, \tag{2}$$

where α represents an intercept, β represents a slope (i.e., the predicted change in Y given a one unit change in RIE), and ε is a residual term. If we substitute the right hand side of Eq. (1) for RIE in the above model and expand the products, it can be seen that the investigator is actually modeling:

$$Y = \alpha + \beta E_{\mathrm{W}} - \beta E_{\mathrm{B}} + \varepsilon. \tag{3}$$

Eq. (3) reveals a model that is restrictive in form. It asserts that the impact of the two implicit evaluations on the criterion is additive and that the effects of each implicit evaluation (as reflected by β) are equal in magnitude but opposite

[1] Written descriptions of the theoretical construct suggest that researchers do view the IAT as a measuring a difference in two implicit evaluations. As examples, Greenwald et al. (1998) often refer to attitudinal *differences* when summarizing IAT results. On page 1476, they state, "In Experiment 2, the expected correlation was in the relationship of an IAT measure of *attitude difference* between Korean and Japanese ethnicities and subjects' self-described ethnic identities ..." (italics added) On page 1474, they state, "The data of Experiment 3... clearly revealed patterns consistent with the expectation that White subjects would display an implicit attitude *difference* between the Black and White racial categories." (italics added) On page 1479, they state, "Findings of three experiments consistently confirmed the usefulness of the IAT (implicit association test) for assessing *differences in evaluative associations* between pairs of semantic or social categories." (italics added) In addition, when IAT researchers compare the IAT to single-attitude explicit or other implicit measures, they first compute the simple difference score for two single-attitude explicit measures (e.g., by subtracting the explicit attitude towards Blacks from the explicit attitude towards Whites). Most typically, these difference scores are then correlated with IAT scores (see Greenwald et al., 1998).

in sign. To the extent that this causal structure is not operating, Eq. (2) provides suboptimal prediction of the criterion, Y, and can yield inappropriate inferences.

To use the above formulated IAT in a criterion prediction context, the theorist must be willing to accept the causal structure it embraces, even when others are plausible. Consider as an example investigators who want to predict racist actions in a sample of White participants as a function of their implicit evaluations of Whites and their implicit evaluations of Blacks. It may be that racist actions of Whites are driven more by implicit evaluations of Blacks (E_B) than by implicit evaluations of Whites (E_W). This suggests that the regression coefficient for E_B will be larger than the regression coefficient for E_W. In such a circumstance, the regression coefficient one obtains from regressing racist actions onto a valid measure of racial preference as conceptualized by the IAT will be an intermediate value between the (relatively large) coefficient one would have obtained from a valid measure of E_B and the (relatively small) coefficient one would have obtained from a valid measure of E_W. As another example, suppose that negative evaluations of Blacks lead to more racist actions in a sample of Whites, the more positively participants evaluate other Whites. This implies an interaction between E_W and E_B. This causal dynamic cannot be detected with the above formulated IAT.

In sum, because the IAT is designed to measure relative implicit evaluations at the conceptual level, it invokes a restrictive causal model that may not be appropriate for many research enterprises. We explore the ramifications of these limitations in the two studies that follow.

The observed level: Relative measurement models

In addition to embracing a relative construct at the conceptual level, IAT researchers embrace a relative construct at the observable level. However, this construct is quite different from the one at the conceptual level. The relative construct at the observed level does not focus on the difference between two implicit evaluations. Rather it focuses on the difference between two *response latencies*, each of which is influenced by the same underlying relative implicit response. To explore the implications of this for the IAT measurement model, we first review the data collection and scoring methods surrounding the IAT. We focus our attention only on the "traditional" IAT methodology. New variations of the IAT and new scoring methods recently have been introduced (Greenwald, Nosek, & Banaji, 2003; Nosek, Greenwald, & Banaji, 2005), but these revisions do not address our critique. A discussion of these new scoring procedures is in Appendix A.

IAT methodology

Measurement strategy

Participants taking the IAT are asked to correctly categorize four types of stimuli that are presented one at a time on a computer screen. Stimuli are categorized into one of two categories via a key press. With the race IAT, as in all IAT, two of the stimuli types are related to the two attitude objects that are of interest and two of the stimuli types are related to the endpoints of the evaluative dimension upon which these objects are being evaluated. Thus, a version of the race IAT that is designed to assess the bipolar evaluation (positive to negative) of Whites versus Blacks would have stimuli related to positive (positively valenced words), negative (negatively valenced words), Blacks (pictures of Black faces) and Whites (pictures of White faces).

Relative implicit evaluations are assessed by measuring how long it takes for participants' to assign the stimuli to the experimenter-defined categories. The categories are created by pairing each attitude object with each of the two evaluative anchors. For example, one category pair might be "Whites or Positive" and another might be "Blacks or Negative." Two different categorization tasks are given. One task, called the "compatible judgment," is thought to be easy for people who are high on the construct of interest and the other, called the "incompatible judgment," is thought to be hard for people who are high on the construct of interest. If one were measuring the implicit preference for Whites over Blacks using the race IAT just described, the compatible task would be the one that asks participants to categorize the stimuli into one of the two categories "Whites and Positive" versus "Blacks and Negative" and the incompatible task would be the one asking participants to use the categories "Whites or Negative" versus "Blacks or Positive." For example, a Black face might be shown on a computer screen and, for the compatible task, the participant indicates by pressing one of two keys whether the presented stimulus is in the category "Whites or Positive" or in the category "Black or Negative." The logic is that it is easier for someone who prefers Whites to Blacks to categorize Whites with Positive and Blacks with Negative than it is to categorize Whites with Negative and Blacks with Positive. This should result in faster response times for the compatible task as opposed to the incompatible task.

Response latencies for these two tasks are measured in milliseconds, but scores typically are Winsorized to have endpoints of 300 and 3000 ms and then log transformed. The IAT score is then computed by subtracting the transformed compatible score from the transformed incompatible score. With the race IAT, this value is thought to represent the magnitude of the implicit preference for Whites over Blacks. If, averaging across trials, responses to the compatible task are faster than responses to the incompatible task, this is interpreted as an implicit evaluative preference for Whites over Blacks. If, averaging across trials, responses to the compatible task are slower than responses to the incompatible task, then this is interpreted as an implicit evaluative preference for Blacks over Whites (cf. Blanton & Jaccard, in press).

1. [Compatible]

How strong are the associations "Whites and Pleasant" and "Blacks and Unpleasant"?

Extremely strong (300) ____:____:____:____:____:____:____ (3000) Extremely weak

2. [Incompatible]

How strong are the associations "Whites and Unpleasant" and "Blacks and Pleasant"?

Extremely strong (300) ____:____:____:____:____:____:____ (3000) Extremely weak

Fig. 1. IAT "item format."

Item format

The IAT measurement strategy is complex. However, we have found that viable interpretations are revealed if one first "decodes" the IAT to show the possible meanings of the incompatible and compatible scores when they are viewed in isolation. For this purpose, we think it is useful to represent the IAT "item format" in the manner shown in Fig. 1. For pedagogical purposes, we depict the two IAT judgments using a semantic differential format that is familiar to social psychologists. We use the endpoints that are the lower and upper limits for reaction time in IAT research (300 milliseconds reflecting a fast response or a strong association and 3000 milliseconds reflecting a slow response or a weak association).

Note in this representation that both IAT tasks assess some aspect of the same relative implicit preference. Persons who have a preference for Whites over Blacks should be *faster* on the compatible task in the race IAT since they presumably have more positive associations with Whites and more negative associations with Blacks. These same people will have *slower* responses on the incompatible task because they have fewer positive associations with Blacks and fewer negative associations with Whites. As this example highlights, the compatible and incompatible task can be viewed as "reverse coded" items that reflect opposing aspects of the *same* relative implicit preference. That is, both capture some comparison between the implicit evaluation of Whites and implicit evaluation of Blacks, but one is easier for those who favor Whites relative to Blacks and the other is harder for those who favor Whites over Blacks.[2] As developed in the next section, the conceptualization of the two opposing IAT judgments as reverse scored indicators of the same construct leads to a straightforward psychometric model for the IAT; one that can be empirically tested.

Psychometric models

Single-factor IAT model

The first model we consider has a venerable history in psychometrics. This model assumes that there is a single latent and unobservable IAT variable (the implicit relative preference for Whites relative to Blacks) that influences the two observed variables (the compatible and incompatible reaction times).[3] A common representation in the field of psychometrics is to view the observed scores of psychological measures as being a linear function of the latent construct of interest. With the two observed scores of the IAT, this can be expressed algebraically as:

$$\mathrm{IRT} = \alpha_1 + \lambda_1 \mathrm{RIE} + \varepsilon_1,$$
$$\mathrm{CRT} = \alpha_2 + \lambda_2 \mathrm{RIE} + \varepsilon_2,$$

where IRT is a person's reaction time for the incompatible IAT task, CRT is a person's reaction time for the compatible IAT task, RIE is a person's true (unobservable) score on the relative implicit evaluation, the α are intercepts, the λ are slopes or regression coefficients, and the ε represent random measurement error.

Fig. 2 presents a causal model reflecting this measurement conceptualization for the case in which one is trying to predict a psychological outcome, Y, from an implicit relative evaluation. The researcher has three measures: Two measures that reflect the implicit relative evaluation (i.e., response latencies for the compatible and incompatible categorization tasks) and one measure of the criterion (Y). Following traditional conventions, the rectangles in Fig. 2 indicate observed measures and the circles indicate latent variables that represent either the true underlying construct in question or a residual term. The λs represent path coefficients from the latent variables to the observed measures, the βs represent path coefficients for latent variables and the εs represent the residuals.

As noted, IAT researchers estimate the true relative implicit evaluation in practice by computing a difference between the (transformed) compatible and incompatible reaction times. Computation of a simple difference score carries with it certain analytic assumptions (see Carver, 1989; Cronbach & Furby, 1970; Griffin, Murray, & Gonzalez, 1999; Johns, 1981). If the psychometric model in

[2] By the same token, individuals who favor Blacks over Whites will find the incompatible task easier (i.e., they will respond faster to it) and the compatible task harder (i.e., they will respond to it slower).

[3] In our analyses of IAT psychometric models, it does not matter whether one views the *conceptual* IAT score as an unweighted difference, a weighted difference, as a ratio or in any other way in which the IAT might assess a "relative" construct. Our analysis of psychometric model applies to any theoretical construct that might be represented by computing a difference between the two *observed* IAT responses.

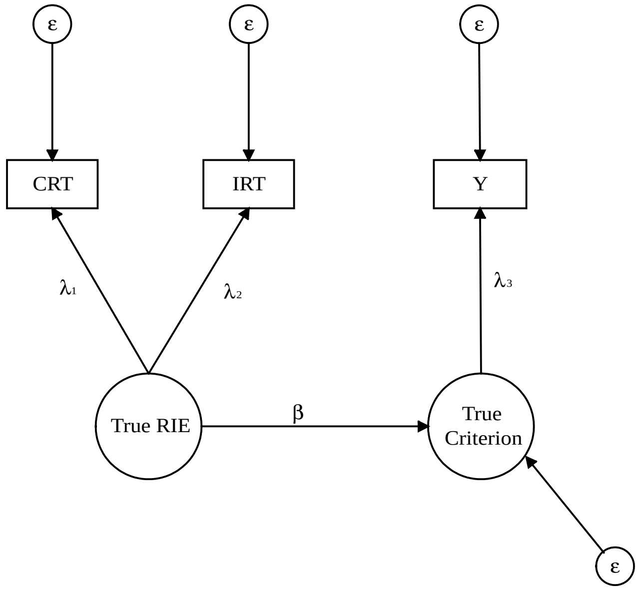

Fig. 2. Single-factor IAT model.

Fig. 2 applies to the IAT, computation of a difference score assumes that λ_1 and λ_2 are equal in value and opposite in sign. Viewed from a more traditional psychometric perspective, calculating a difference score in this context is mathematically equivalent to treating the incompatible and compatible reaction times as positively loading and negatively loading "items" of the same psychological factor. Note how this single-factor conceptualization of the IAT is consistent with the conceptualization that follows from the IAT item format, as shown in Fig. 1.

Unfortunately, the single-factor model is not empirically defensible. If one were to collect multiple measures of both the compatible and incompatible reaction times and pursue formal testing of the model in Fig. 2, the results for virtually any IAT would fail to uncover empirical support for the assumption that $\lambda_1 = -\lambda_2$. This is because the correlation between the compatible and incompatible measures for the IAT typically is *positive*, whereas the above psychometric model predicts that they should be *negative* (because λ_1 and λ_2 are assumed to be opposite signed). In our own data sets, we have found that the correlation between the two IAT response latencies range anywhere from 0.40 to 0.60.

Single-factor IAT model with systematic error

There are other models that are theoretically defensible, that support the decision to compute a difference score and that can tolerate a positive correlation between IRT and CRT. These models assume that any positive correlation between the two IAT component scores is the result of one or more common sources of systematic method variance. It is reasonable to consider models of this kind, because it is well known that IAT component scores are influenced by at least one systematic source or error. More specifically, because the two IAT judgments are both reaction time measures, they are each influenced by a person's ability to respond quickly to cognitive tasks. We refer to this ability as a person's *general processing speed* and we view it as a psychological resource that may be imperfectly representative of such general attributes as cognitive efficiency, general intelligence and attention span. It also may reflect such skills as finger dexterity and hand-eye coordination and more transitory states such as mood, drug use and the like. The influence of this common source of method variance on the two component judgments of the IAT is shown in the psychometric model in Fig. 3.

The fact that compatible and incompatible response latencies show a moderate to strong positive correlation at the observed level suggests that general processing speed exerts a greater systematic influence on the two component scores of the IAT than do the implicit relative evaluations that these responses are thought to measure. In studies using the IAT, one must therefore develop strategies to eliminate or control for this source of systematic measurement error.

We have found that many researchers appreciate the need to control for general processing speed in the IAT. In fact, many believe such the IAT was designed to address this concern. When we conducted an informal poll of researchers and asked them why they think a difference score has been built into the IAT measure, many spontaneously mentioned a belief that analyses need to focus on the difference scores in order to eliminate large individual differences in general reaction times. There is a long history in cognitive psychology for computing "facilitation scores" by subtracting one reaction time measure from another. This tradition might cause some to assume that processing speed somehow is "factored out" by the calculation of a difference score for the IAT. Although differencing can

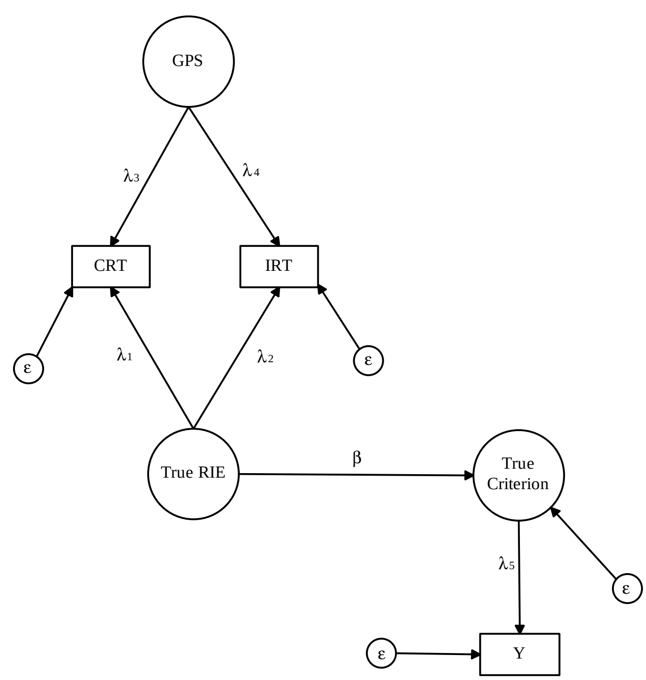

Fig. 3. Single-factor IAT model with systematic error.

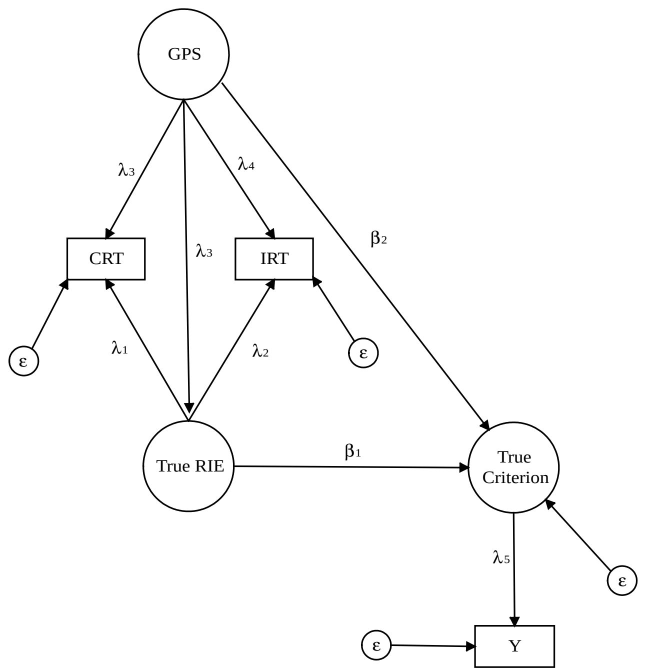

Fig. 4. Single-Factor IAT model with meaningful and nuisance effects of processing speed.

eliminate a common source of systematic error that influences two observed scores, the conditions needed for this to occur with the IAT are complex and may be difficult to justify on theoretical grounds (see Appendix A for a fuller discussion).

Single-factor IAT model with meaningful and nuisance effects of processing speed

An alternative strategy for dealing with processing speed in the IAT is not to trust differencing (or some other arbitrary scoring algorithm) to remove general processing speed. Rather, one can directly model the effect of such confounds in the data. Fig. 4 presents an IAT psychometric model that is general in nature and that reflects these underlying dynamics.[4] The paths from the latent variable of processing speed to the observed measures of the IAT (λ_3 and λ_4) are intended to reflect systematic error variance in the IAT measures due to processing speed. The model also allows for meaningful effects of general processing speed on the latent relative implicit evaluation (β_3) as well. This path recognizes that variables reflected by general processing speed (e.g., intelligence) may sometimes bear meaningful rather than artifactual relations to implicit attitudes, such as the where intelligence is related to prejudice. The model also allows for direct influences of general processing speed on the true criterion of interest (β_2), a feature that neither difference coding nor other scoring algorithms offered by Greenwald et al. will permit (see Appendix A).

Study 1

Study 1 had two purposes. First, we tested the viability of the relative-preference causal model. Second, we tested the appropriateness of the above measurement models that can justify computation of a difference score. In this study, we decided to focus attention on a situation in which the IAT has shown strong prediction in the past. We did this because such a test should provide the most favorable analysis of the IAT's value as a research tool. We thus performed a secondary data analysis on a study designed to replicate a finding in Nosek, Banaji, and Greenwald (2002), Gonzales (2002). In our analyses, we tested the theory that "implicit preference for math over arts leads to explicit math identification." We admit that the theory we tested was mundane and certainly less interesting than the primary concerns investigated by Nosek, Greenwald and Banaji in their original paper. However, our goal was not to test an interesting theory or a counter-intuitive relationship, but rather, to test an obvious theory and a strong relationship. In this way, we could conduct a liberal test of the causal model built into the IAT at the conceptual level and the measurement models built into the IAT at the observed level.

Study details

The total sample consisted of 215 undergraduate psychology students who participated for course credit in a larger study of math and arts attitudes, gender stereotyping and gender identification. Following IAT conventions, we eliminated 13 participants who had more than 15% error rates. The final sample thus consisted of 202 participants (117 female, 85 male), ages 17–24 (mean = 18.47, $SD = 1.24$). The study followed Nosek et al. (2002) and had participants categorize experimental stimuli related to math (e.g., calculus, algebra, and geometry), the arts (e.g., symphony, drama, and literature), pleasant (e.g., love, rainbow, and heaven), and unpleasant (e.g., death, torture, and hatred) across 80 experimental trials. In one judgment task (involving 40 of the trials), participants were instructed to categorize the above stimuli with the super-ordinate categories "Math or Pleasant" versus "Arts or Unpleasant" and in another judgment task (involving the other 40 trials) the super-ordinate categories "Math or Unpleasant" versus "Arts or Pleasant."

Following the procedures in Greenwald et al. (1998), we eliminated participants who had more than 15% error rates, we Winsorized values to range from 300 to 3000, and we applied log transforms.[5] We also eliminated the first 4 trials to provide participants with time to familiarize themselves with the response set. The resulting scores ran-

[4] The model in Fig. 4 is underidentified for statistical analysis. The figure is only intended to illustrate the psychometric model on a conceptual level.

[5] Our data coding also differed from Gonzales, as she did not apply the logarithm transformation in her analyses. We did this in the current study to be consistent with Nosek et al. (2002).

ged from 5.7 to 8.01, with larger values representing longer response latencies. For ease of presentation, we recoded these values to represent speed of response by subtracting them from the value 8.01. We labeled the first judgment $M + A-$, because it simultaneously assessed the tendency to associate math with positively valenced words and arts with negatively valenced words. Scores on this variable were coded so that higher values implied a preference for math over arts. We labeled the second judgment as $M - A+$, because it simultaneously assessed the tendency to associate math with negatively valenced words and arts with positively valenced words. Scores on this variable were coded so that higher values implied a preference for arts over math. To replicate past scoring strategies, we computed a difference between the $\mathrm{M} + \mathrm{A}-$ and $\mathrm{M} - \mathrm{A}+$. Positive scores on this difference score traditionally would be interpreted as showing a degree of "preference for math over arts" and negative scores on this difference score traditionally would be interpreted as showing a degree of "preference for arts over math."

For the model testing strategies we describe later, it is necessary to have multiple measures of $\mathrm{M} + \mathrm{A}-$ and $\mathrm{M} - \mathrm{A}+$. We adopted a "random blocking" strategy to accomplish this. Because each type of judgment is replicated on 36 trials (the 40 original trials minus the 4 practice trials), one can create four indicators of a given type by randomly dividing the 36 trials into 4 subsets of 9 trials each and then calculating the average response time for each block of 9 trials. Although the score for a given block will be less reliable than the total score across all 36 trials, the analytic methods we propose can adjust for this. If desired, a researcher can increase the number of trials that comprise a sub-block by generating fewer indicators (e.g., three indicators based on subsets of 12 trials each) or by increasing the overall number of trials in the study (e.g., collect data on 60 or 80 trials for that type of judgment).

To test a measurement model that incorporates general processing speed, it also was necessary to develop a measure of it. Our measure was taken from the practice trials of the IAT. As described in Gonzales (2002), all participants were provided a total of 12 practice blocks in which they made simple discriminations to familiarize themselves with the stimulus words for the four category word sets (e.g., "math versus arts" and "pleasant versus unpleasant"). We used the average response latencies from these practice blocks as indicators of each individual's general processing speed. The only constraint was that we discarded the first block in which participants learned the task, and every block after that in which participants were exposed to experimental stimuli for the first time. As a result, we had a total of seven measured indicators of general processing speed. As with the IAT judgments, response times were Winsorized to have a range from 300 to 3000, log transformed and recoded so that higher scores were indicative of faster processing. For our preliminary analyses, we generated a single indicator of general processing speed by averaging the scores for the seven indicators. The more formal model tests used the multiple indicators separately.[6]

In addition to these implicit measures, participants also completed a set of explicit measures. The criterion of interest, math identification, was measured with 3 items (e.g., "I am strongly identified with math"), with scores ranging from 0 to 10. A single indicator of this construct was formed for the preliminary analyses by averaging the responses to the three items. Participants also provided explicit measures of attitudes toward math and attitudes towards the arts. These were multi-item attitude scales that yielded a single overall attitude score for the attitude toward math and a single overall attitude score for the attitude toward the arts. The former scale had a Cronbach alpha of 0.97 and the latter had a coefficient alpha of 0.79 (see Gonzales, 2002, for details). Scores for each attitude could range from -10 (extremely unfavorable) to $+10$ (extremely favorable).

Results

Preliminary statistics

The correlations between the observed measures are presented in Table 1. Several relationships are worth noting. First, the correlation between the measure of the explicit attitude towards math and the explicit attitude towards the arts was small and statistically non-significant, $r(200) = 0.04$. This low correlation suggests that the two explicit attitudes are distinct. This finding should give theorists pause about the idea that these two attitude objects represent "complementary" attitude objects (Greenwald & Farnham, 2000). The utility of maintaining separate constructs at the level of explicit attitudes is reinforced by the tendency for math identification (the criterion) to have a strong association with explicit math attitudes, $r(200) = 0.82$, $p < .001$, but not explicit arts attitudes, $r(200) = -0.11$, ns.

The correlations in Table 1 also suggest that general processing speed is a source of systematic measurement error. The correlation of general processing speed with the $\mathrm{M} - \mathrm{A}+$ implicit response was $r(200) = 0.72$, $p < .001$ and with the $\mathrm{M} + \mathrm{A}-$ response was $r(200) = 0.55$, $p < .001$. The somewhat higher correlation of general pro-

[6] The primary threat to the validity of this measure is that the practice items might be influenced by different cognitive skills from those of the judgments. This could occur because the former involve categorization into one of two single-concept categories and the latter involve categorization into one of two double-concept categories. To address this possibility, we also created an alternative measure of GPS based on IAT judgments. In addition to taking the math/arts attitude IAT, participants also completed IAT for gender stereotyping (women and men paired with math and arts), gender identification (women and men paired with self and other), and math identity. The alternative measure was based on the IAT for gender stereotyping and gender identification. We found that this alternative measure of GPS was strongly related to the one based on practice items, $r = 0.97$. For this reason, we used our initial measure based on the practice items in all analyses reported hencewith.

Table 1
Correlation matrix for implicit and explicit measures

		1	2	3	4	5	6	M	*SD*
Criterion									
1	Math Identification	—						4.50	2.62
Implicit measures									
2	IAT	0.34**	—					−0.18	0.20
3	M + A−	0.22	0.58**	—				1.07	0.20
4	M − A+	−0.13	−0.47**	0.45**	—			1.24	0.18
5	GPS	−0.08	−0.12	0.55**	0.72**	—		1.46	0.14
Explicit measures									
6	Math attitude	0.82**	0.25**	0.19**	−0.06	0.11	—	1.11	4.84
7	Arts attitude	−0.11	−0.26**	−0.06	0.22	0.11	0.04	4.07	2.82

Note. Math identification is an explicit rating that ranges from 0 (low identification) to 10 (high identification). M + A− is the speed of associating math with pleasant and arts with unpleasant and M − A+ is the speed associating arts with pleasant and math with unpleasant. GPS is general processing speed, based on the speed for the practice trials. MP, M − A+ and GPS range from 1 (slow) to 3.31 (fast) and IAT is computed as M + A− minus M − A+. Math and arts attitudes range from −10 (negative evaluation) to +10 (positive evaluation). All double asterisks for .01 significance.

cessing speed with M − A+ suggests that the assumption of equal influence of general processing speed on the two measures may be untenable. Support for the influence of general processing speed also is evident in the positive correlation between the M + A− and M − A+, $r(200) = 0.45$, $p < .001$. As noted earlier, one would normally think these measures should be negatively correlated if they reflect complementary pairs or reverse scored items that can be differenced. This correlation became small and nonsignificant when the measure of general processing speed was partialled out from each of the IAT measures, *partial* $r(199) = 0.09$. One would expect a strong negative correlation between M + A− and M − A+ once they are purified of general processing speed, but the above partial correlation was near zero rather than strongly negative.

The correlations in Table 1 also suggest that general processing speed is not strongly correlated with the criterion of math identification, nor is it strongly correlated with either of the explicit attitude measures. To the extent that the statistical dynamics surrounding the explicit math and arts attitudes are similar to those occurring with implicit attitudes, these results suggest that general processing speed may be of little consequence other than to create unwanted variation in the IAT measures.

The IAT difference score tended to show a higher correlation with the criterion of math identification ($r = 0.33$) than either of its component parts (absolute r of 0.21 and 0.13). It would be imprudent to conclude from this that the higher correlation between the difference score and the criterion reflects anything theoretically special about differencing in terms of modeling the underlying attitude dynamics. The stronger correlation could result because the component measures are more contaminated by general processing speed (and because general processing speed is unrelated to the criterion).

All of the above conclusions must be viewed with caution because they assume reliable and valid measurement and because some hinge on interpretations of null results. Nevertheless, this initial, surface-level analysis suggests potential threats to the math/arts IAT regarding the relative IAT conceptual construct and the measurement models that support a computed difference score. We now turn to more formal methods of model evaluation.

Conceptual level: Causal model analysis

At the conceptual level, the math/arts IAT is designed to assess the relative difference between implicit math attitudes and implicit arts attitudes. In contrast, the explicit attitude measures were designed to assess math and arts attitudes separately. One can combine the explicit ratings, however, to obtain an estimate of the difference between explicit math attitudes and explicit arts attitudes. This can be done by subtracting the explicit arts rating from the explicit math rating. We did exactly this, even though we knew that this computation is a questionable practice.[7] We regressed math identity onto this difference score and found that this new variable accounted for a significant proportion of the variance in the criterion, $R^2 = 0.61$, $F(1, 201) = 319.45$. The regression coefficient indicated that the measure of "explicit preference for math over arts" was a positive predictor of math identification (regression coefficient = 0.38, $p < .001$).

From this result, one might mistakenly conclude that there is something special about the "explicit preference for math over arts" that leads a person to identify with math. A conclusion of this sort is typical of what can be found in IAT manuscripts. Greater insight into the underlying dynamics is obtained by exploring further the causal model that is implied by the difference score. The regression coefficient obtained from testing the influence of this "relative explicit evaluation" on math identification is exactly the same coefficient that one would obtain if one tests a

[7] This coding strategy has a basis in the IAT literature. Researchers commonly compute differences between two explicit measures when assessing the validity of the IAT as a relative measure (Greenwald & Farnham, 2000; Greenwald et al., 1998; McConnell & Leibold, 2001; Swanson, Rudman, & Greenwald, 2001).

structural equation model in which the measures for explicit math attitudes and the measure of explicit arts attitudes are forced to exert equal-but-opposite influences on the criterion. We used the computer program LISREL to evaluate a model that imposed such a constraint and indeed we obtained the same regression coefficient. Importantly, however, the test of the fit of the overall model suggested a poor model fit, indicating a misspecified model (χ^2 (1) = 50.34, $p < .001$; GFI = 0.87; CFI = 0.79, RMSEA = 0.50 with 90% confidence intervals of 0.38–0.62, p value for the test of close fit < .001, standardized RMR = 0.12). As shown below, the lack of fit occurs because of the constraint that the influence for math and arts attitudes must be equal but opposite in sign. This illustrates that a researcher can predict a criterion, even when the analysis conducted is an inaccurate representation of the true causal dynamics surrounding it (see Anderson, 1981). If the "equal but opposite influence" model were correct, we would not have observed the overall ill fit of the model. The fact that we did clearly suggests that one should not orient one's analyses or interpretations around this difference score. It could be that a similar dynamic occurs at the implicit level, i.e., that researchers who use the math/arts IAT are unknowingly imposing poorly fitting causal models in their work when they use it to predict criteria.

To gain more insights into the underlying causal model, we removed the constraint of equal and opposing main effects so that the two predictors could have differential influence on the criterion. This model was "just identified," so perfect model fit was assured. Importantly, the estimated values of the path coefficients (i.e., the regression coefficients) revealed why the difference score approach yielded significant prediction despite its inaccurate representation of the underlying causal structure. There was a large and positive path coefficient for math attitude, $B = 0.45$, $p < .001$ but a relatively small and negative influence of arts attitude, $B = -.14$, $p < .001$. As a result of these two counter-veiling influences, the difference score provided reasonable prediction of math identity. However, the opposing influences were not of equal magnitude (as indicated by the poor fit of the first model relative to the second model). Math identification thus did not seem to derive from a "preference of math over arts," as is implied when one uses a predictor variable that is meant to assess exactly this conceptual construct. Instead, math identity seemed to be driven mostly by the attitude toward math, with a small contribution from the attitude toward the arts.

Observed level: Measurement model analysis

We turn now to an analysis of the observed IAT difference score, which is computed by subtracting M − A+ from M + A−. Tests of the IAT measurement model using structural equation modeling are facilitated by the use of multiple indicators for the underlying latent variables. To this end, we used the seven indicators of general processing speed and the four measures of M − A+ and M + A− discussed earlier.[8] The analyses were implemented using LISREL 8.5 in conjunction with its non-linear constraint feature (Jaccard & Wan, 1996).

The first measurement model that we tested assumed that the covariation of the eight IAT *observed* measures could be explained by a single latent variable representing the *conceptual* relative preference for math over arts. To justify the use of a difference score, the model was constrained so that M − A+ measures would have loadings that were equal but opposite in sign to loadings for M + A− (see Fig. 2). The metric of the latent variable was defined by constraining its variance to 1.0, i.e., treating it as if it were in standardized metric. This model ignored potential influences of general processing speed on the measures. Instead, it simply assumed that the IAT judgments were "pure" indices of implicit relative evaluations. Not surprisingly, the model yielded poor fit (χ^2 (24) = 922.8, $p < .001$; GFI = 0.32; CFI = 0.04, RMSEA = 0.59 with 90% confidence intervals = 0.57–0.61; p value for close fit < .001; standardized RMR = 0.44). This indicates that a model following traditional psychometric traditions is untenable. This also was true for a liberalized single factor model that imposed no constraints on the factor loadings (χ^2 (20) = 337.6, $p < .001$; GFI = 0.61; CFI = 0.66, RMSEA = 0.35 with 90% confidence intervals = 0.33–0.38; p value for close fit < .001; standardized RMR = 0.16).

Although the conventional model was not a reasonable representation of the mechanisms that underlie the data, it is possible that the theoretical essence of the single-factor model would apply if general processing speed is taken into account. To explore this, we tested a model that assumed a single factor relative implicit evaluation, with contaminated measurement. The eight observed measures were modeled to be a function of a single IAT latent construct as well as a second latent variable representing general processing speed. The model is presented in Fig. 5. To avoid clutter, Fig. 5 omits random measurement error for each of the observed measures as well as the correlation between the two latent variables, although these parameters were estimated. The model is under-identified, so constraints were introduced to achieve model identification. The first set of constraints were those invoked previously and that are inherent to the IAT framework, namely equal-but-opposite loadings for the respective paths from the latent implicit attitude variable to the observed latency measures ($\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$ and $\lambda_5 = \lambda_6 = \lambda_7 = \lambda_8$, with $\lambda_5 = -\lambda_1$, $\lambda_6 = -\lambda_2$, $\lambda_7 = -\lambda_3$, $\lambda_8 = -\lambda_4$). The second set of con-

[8] We conducted preliminary analyses on the viability of the seven processing speed indicators based on the practice trials as meaningful representations of a single underlying factor of GPS. The test of a single factor model yielded indices that were consistent with good model fit. In addition, the standardized path coefficients from the standardized latent variable to the standardized observed measures tended to be high and reasonable in magnitude. Thus, ill fit of the models reported later cannot be attributed to ill fit in this aspect of the model.

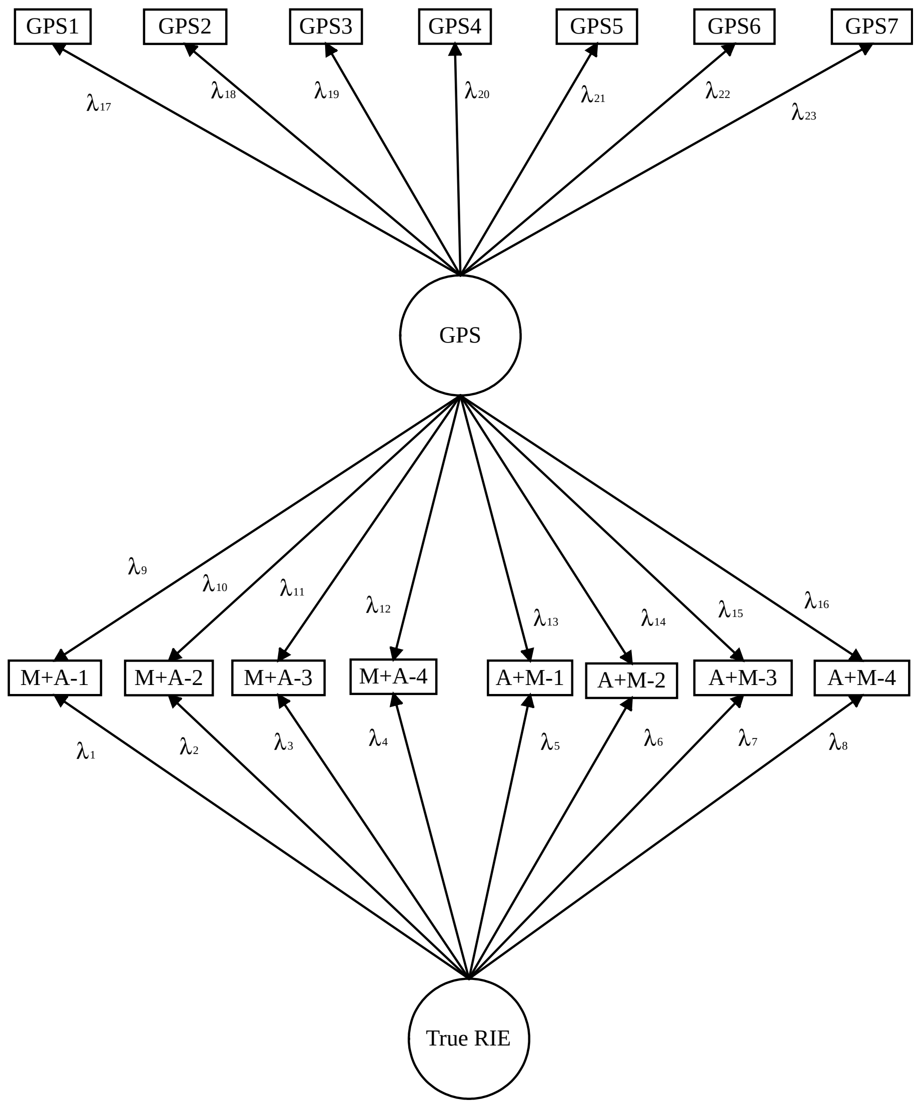

Fig. 5. Math/arts IAT measurement model with general processing speed.

straints was based on the assumption that general processing speed has an equal influence on each of the response latencies ($\lambda_9 = \lambda_{10} = \lambda_{11} = \lambda_{12} = \lambda_{13} = \lambda_{14} = \lambda_{15} = \lambda_{16}$; see Appendix A for a discussion of this assumption.) As before, the variances of the latent variables were constrained to be 1.0. This model also yielded an unsatisfactory fit to the data (χ^2 (95) = 247.7, $p < .001$; GFI = 0.81; CFI = 0.91, RMSEA = 0.11 with 90% confidence intervals = 0.10–0.13, p value for close fit < .001; standardized RMR = 0.07).

We next estimated a liberalized version of the two factor model where we relaxed the constraint of equal influence of general processing speed on the observed IAT measures. This model constrained the influence of general processing speed to be equal for M + A− ($\lambda_9 = \lambda_{10} = \lambda_{11} = \lambda_{12}$) and M-A+($\lambda_{13} = \lambda_{14} = \lambda_{15} = \lambda_{16}$), but not across the two classes of measures (i.e. it was not the case that λ_9, λ_{10}, λ_{11} and λ_{12} had to equal λ_{13}, λ_{14}, λ_{15}, and λ_{16}). In addition, the model permitted all of the parameters from λ_1 through λ_8 to be free to vary. The model yielded borderline adequate fit for some of the fit indices but not others (χ^2 (87) = 188.6, $p < .001$; GFI = 0.87; CFI = 0.94, RMSEA = 0.08 with 90% confidence intervals = 0.06–0.09; p value for close fit < .001; standardized RMR = 0.05). The standardized path coefficients (factor loadings) from the latent IAT to the M + A− indicators were reasonable in magnitude ($\lambda_1 = 0.62$, $\lambda_2 = 0.78$, $\lambda_3 = 0.62$, $\lambda_4 = 0.66$), but the corresponding coefficients for M-A+ (which are predicted by the IAT to be opposite in sign to the M+A- indicators) were much lower in magnitude ($\lambda_5 = -0.16$, $\lambda_6 = -0.12$, $\lambda_7 = -0.14$, $\lambda_8 = -0.15$). At best, this model suggests that the IAT measures load differently on the underlying implicit attitude, questioning the traditional IAT computation methods. Only two of the six fit indices suggested adequate model fit, so a more rigorous interpretation is to reject the viability of even this liberalized version of the psychometric model underlying the IAT.

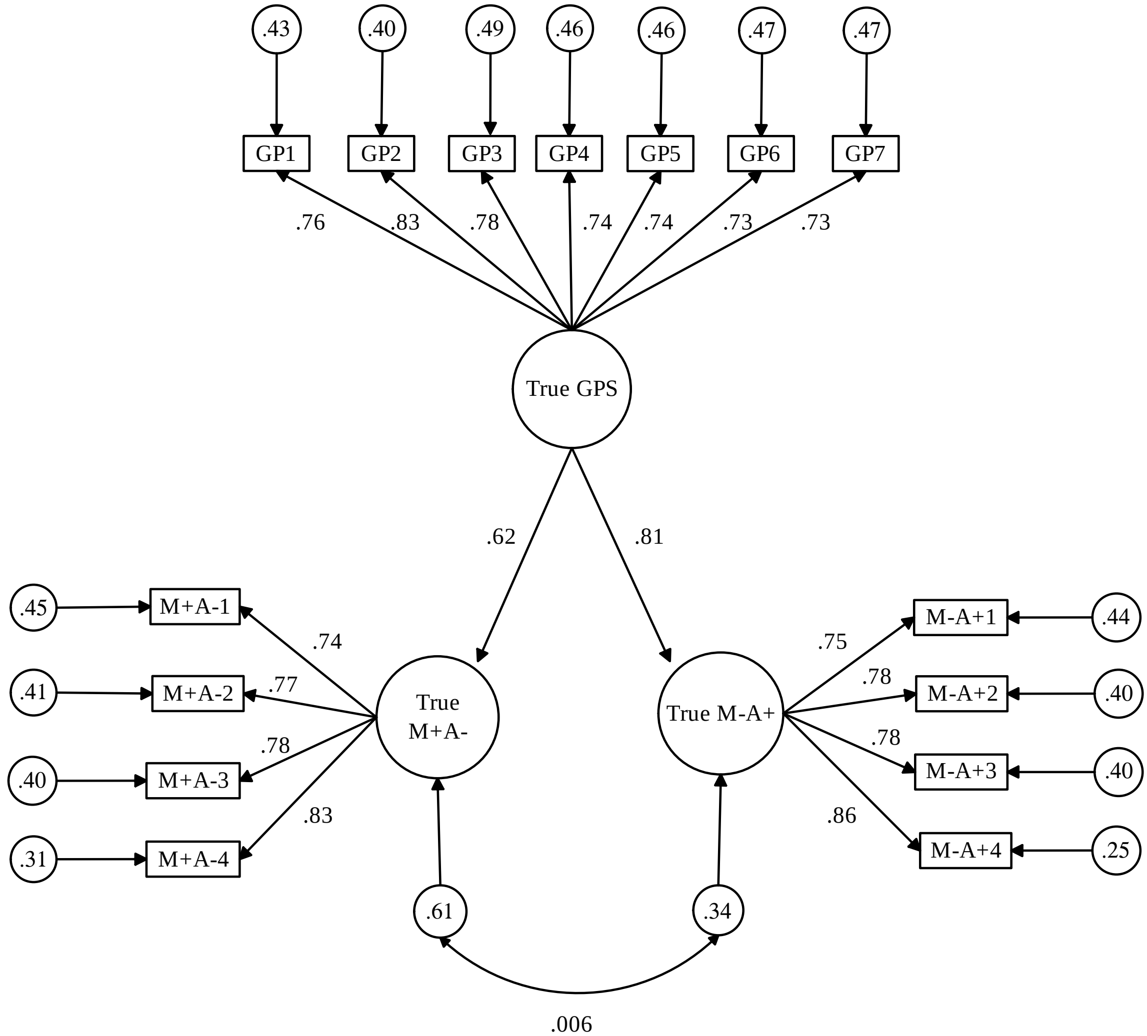

Fig. 6. Partial correlation between M + A– and M – A+ partialling out general processing speed.

Exploratory two-factor model with processing speed

Fit diagnostics in the above analyses suggested that a differently structured two-factor model could best account for the data. For this model, the M + A– reaction times are influenced by one implicit attitude construct and the M – A+ reaction times are influenced by another implicit attitude construct. This model is shown in the upper portion of Fig. 6. One would expect from the IAT scoring methods that these constructs would be negatively correlated, but in our analyses, the correlation was 0.51, $p < .01$. This positive correlation could occur because these two constructs also share a common influence of general processing speed. The model was therefore altered so as to covary out general processing speed from the two latent attitude variables, with the partial correlation being reflected in the correlated residuals in Fig. 6. The model yielded good fit to the data (χ^2 (87) = 134.3, $p < .001$; GFI = 0.92; CFI = 0.97, RMSEA = 0.05 with 90% confidence intervals = 0.03–0.07; p value for close fit < .401; standardized RMR = 0.046). There were no offending estimates and no notable modification indices. The standardized parameter estimates are presented in Fig. 6. The estimated partial correlation between the two IAT constructs after controlling for processing speed was not negative, as one might predict, but instead it was near zero, $r = 0.006$ (ns).

We next used this model as a basis for a criterion-prediction model that predicted the math identification outcome (a latent variable with three indicators) from the two attitude constructs with general processing speed assuming the role of a covariate (see Fig. 7). The model yielded good fit to the data (χ^2 (129) = 181.1, $p < .002$; GFI = 0.91; CFI = 0.98, RMSEA = 0.04 with 90% confidence intervals = 0.03–0.06; p value for close fit < .79; standardized RMR = 0.05). The standardized regression coefficients for the latent variables were 0.32 for the latent M + A– ($p < .002$), −0.72 for the latent M – A+. ($p < .001$) and 0.50 for general processing speed ($p < .001$).[9] It appears from the coefficients that explicit math identification is influenced primarily by the latent implicit M – A+, such

[9] All significance tests of parameters were replicated using bootstrapping and the results yielded comparable conclusions throughout all of our analyses.

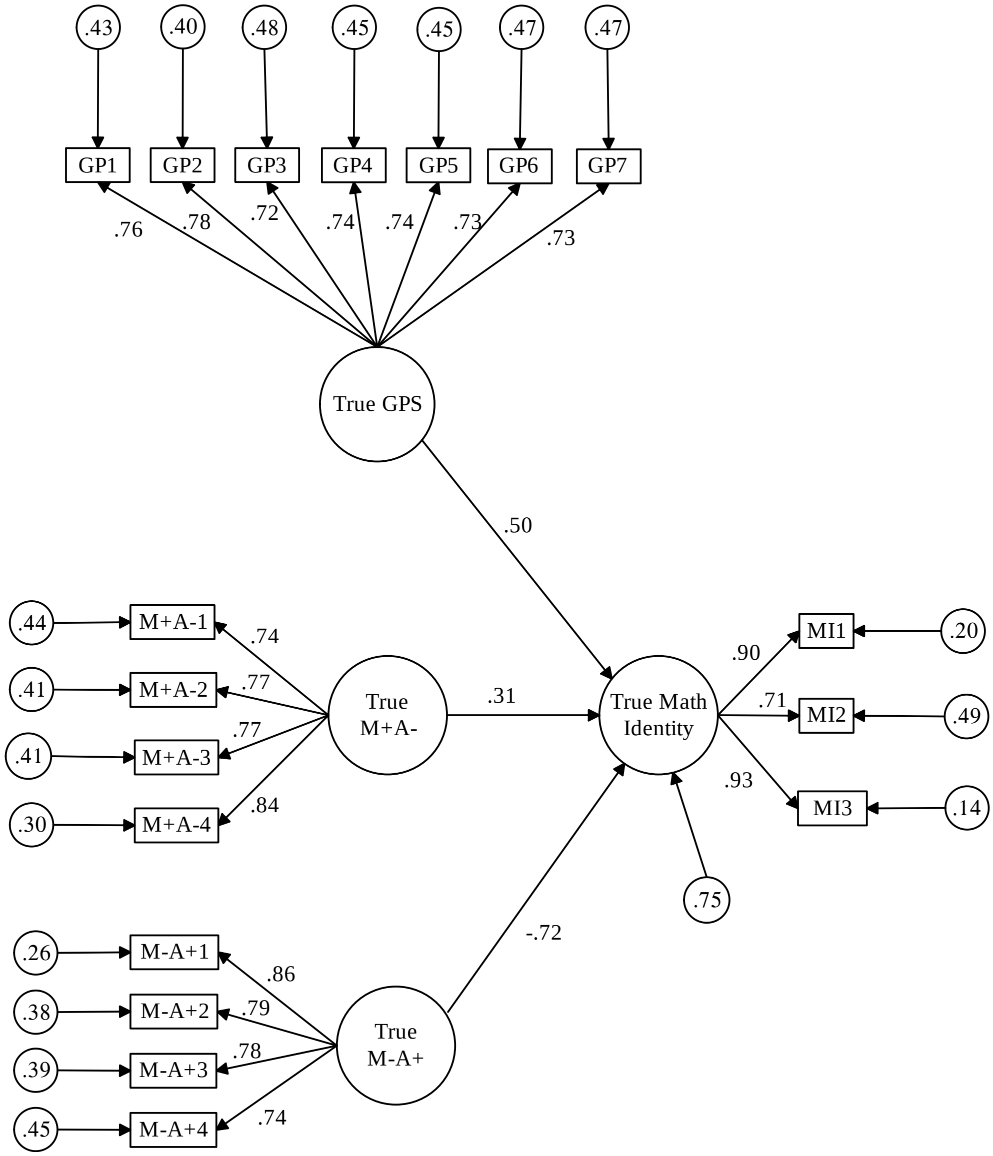

Fig. 7. Latent variable regression of $M + A-$, $M - A+$, and general processing speed on math identity. Correlations between exogenous variables are not shown but were estimated.

that explicit math identification is higher for people who are slow to associate math with negative words and arts with positive words.[10] It is likely that this effect reflects the causal influence of negative math associations, not the effect of positive arts associations, but, one cannot be sure because the two associations are confounded in the IAT measurement strategy. There also was a significant $M + A-$ effect, but it was statistically significantly smaller than the effect for $M - A+$. These findings suggest that an equal weight differencing model for predicting math identification from these constructs is not appropriate. Finally, the results indicate that those with quicker reaction times have a stronger identification with math. This suggests that processing speed is reflective of general ability-related individual differences that influence a person's academic orientation independent of implicit attitudes.

Discussion

Study 1 was designed to replicate a documented IAT phenomenon for the purpose of testing the appropriateness of (a) the IAT causal model that can justify interest in "relative implicit attitudes" at the conceptual level and (b) IAT measurement models that can justify computing a difference between the two observed IAT scores. Our findings raised questions in both respects.

With regards to the IAT causal model, the assumption of equal-but-opposite influences of math and art attitudes was violated when analyses focused on explicit measures. This suggests caution when assuming such a causal struc-

[10] A nested chi square test of the difference between the two unstandardized coefficients for the IAT predictors, adjusted to have a common sign, was statistically significant ($\chi^2 = 6.60$, df = 1, $p < .01$).

ture with implicit measures. Although it is possible that the causal dynamics are different at the level of implicit attitudes, we know of no theory that would predict this to be the case. Our data suggest that researchers should use caution and ensure that they are studying a phenomenon that fits the causal model of equal but opposite influences of the two implicit attitudes.

Concerns also were raised with regards to the single factor IAT measurement model and variants of it. The lack of fit of for the single-latent models—even in models that took into account general processing speed—suggests that the M + A− and M − A+ judgments do not assess the same psychological construct. Our model tests suggested an alternative measurement model that treats the "compatible" and "incompatible" IAT judgments as distinct psychological constructs that can have different influences on psychological criteria. These differential influences argue against calculating a simple difference score between the constructs when predicting a criterion, as traditional IAT practice would dictate. Such a practice can result in application of a misspecified model.

An obvious question is whether the two-factor structure of the IAT that we isolated replicates. It does. In a recent study (Christie, Blanton, & Jaccard, 2005), we used a sample of 223 students to investigate the race IAT. We also have investigated the self-esteem IAT (associating Self versus Other with Pleasant versus Unpleasant) by reanalyzing the original data from Greenwald and Farnham (2000, Study 1).[11] And, we have investigated the gender-stereotyping IAT (associating women versus men with math versus arts) and the gender identification IAT (associating women versus men with self versus other) by analyzing other measures administered to the current sample (Gonzales, 2002). In every instance, we observed poor model fit when we imposed models that assumed that the IAT measured a single implicit relative attitude construct. All IAT measures we have examined instead have the two-factor factor structure observed in Study 1.

Our models not only suggest that IAT research should move forward using the separate components of the IAT as predictors, but that one also must confront the fact that general processing speed is a major source of variation in these component scores. Sometimes this variation will be artifactual and hence, it should be controlled for. But other times, general processing speed will reflect individual difference variables that are a meaningful source of variation in implicit attitudes. The IAT as traditionally applied is limited because it demands that researchers treat all of the individual difference variables that are manifest as general processing speed as artifactual (see Appendix A).

Why the model misspecification?

We believe that a limitation of the IAT as revealed in Study 1 has to do with the IAT item format, shown in Fig. 1. Note, that each "question" in the IAT has what psychometricians call a "double barreled format." That is, participants reveal their standing on two psychological constructs via a single response. In psychometric scale construction using explicit measures, such items are avoided because they often cannot represent individuals' true evaluations. For example, individuals who have positive evaluations of both Whites and Blacks will find it difficult to represent their attitudes if they are asked to agree or disagree with the statement "Whites are good and Blacks are bad." To the extent that explicit attitudes towards Whites and Blacks are not (a) biploar in nature and (b) strongly negatively correlated with one another, then it is difficult to make sense of responses to such statements. For the race IAT, response latencies on just the compatible task are influenced by two associations, (1) the association between Whites and positive concepts and (2) the association between Blacks and negative concepts. This IAT 'item" is thus double-barreled, in the sense it is influenced by two distinct association strengths. If the two associations being measured with this one task are strongly and negatively correlated, then assessing them with a single task is not problematic. If their correlation is weak, then this measurement strategy is questionable. Similar arguments can be made for the incompatible task. Adding an additional level of complexity, if the two associations measured in the double-barreled compatible task and the two associations measured in the double-barreled incompatible tasks do not have strong and meaningful inter-correlations, then one would not expect the two IAT response latencies to load on to a single latent factor.[12]

Where from here?

We believe that a fruitful area for new research will be to design new measures that do not suffer from the ambiguities inherent in the IAT. In this spirit, Study 2 introduces a new implicit association test that we call the *Simple Association Test* (SAT). This measure seeks to separate out the association types that are arbitrarily grouped together in the IAT. It is a single-target IAT in that it seeks to measure the implicit evaluations of a single attitude object rather than a relative preference. A number of other single-target implicit attitude measures have been developed recently (De Houwer, 2003; Fazio, Jackson, Dunton, & Williams, 1995; Karpinski & Steinman, in press; Nosek & Banaji, 2001; Schnabel et al., 2005). Although these new measures

[11] We thank Tony Greenwald for making these data available to us.

[12] With the race IAT, the correlation structure that will justify the double-barreled format and that will lead to a single-factor solution is one in which the positive associations for Blacks are strongly and negatively correlated with the positive associations for Whites, the negative associations for Blacks are strongly and negatively correlated with the negative associations for Whites, and the positive and negative associations for the same attitude object are strongly and negatively correlated. If this correlation structure is not operating, then one should be suspect of the use of the IAT. No studies to date have tested if such a structure exists for any current IAT measures.

have the potential to test a wider range of causal models than can be examined with traditional versions of the IAT, researchers who have developed the measures have not tested whether their computational strategies are supported by empirically grounded measurement models. Rather, test designers simply perform operations on their data that seem intuitively meaningful or that have precedence in the field. Study 2 illustrates a strategy for developing implicit measures that have defensible measurement models. In the process, the results provide insights into why the two IAT component judgments did not load on a single attitude construct in Study 1.

The control-category IAT: Making a relative measure a non-relative measure

To highlight a critical difference between the SAT and the traditional IAT, we first discuss a strategy that researchers have used in the past, in an apparent effort to make the traditional IAT less relative in nature. This is the incorporation of a "control category" that is not expected to have a causal influence on a criterion. As an example, Swanson et al. (2001) investigated smoking using an IAT that assessed the relative implicit attitudes for "smoking" versus "sweets." Intuitively, "sweets" seems to be an attitude object that is irrelevant to smoking, and so one might conclude that it is inconsequential to incorporate implicit evaluations of sweets into a research program that is focused on predicting smoking criteria. This is not necessarily the case. If one regresses a criterion onto two implicit evaluations separately, one obtains:

$$Y = \alpha + \beta_1\, E_{\mathrm{SMOKING}} + \beta_2\, E_{\mathrm{SWEETS}} + \varepsilon \qquad (4)$$

If it is the case that the implicit evaluation of sweets is not related to the criterion, then β_2 is zero and the equation becomes

$$Y = \alpha + \beta_1\, E_{\mathrm{SMOKING}} + \varepsilon$$

and the criterion is a simple linear function of the implicit attitude toward smoking. Recall, however, that the IAT measures a difference in smoking and sweets attitudes, so one makes the assumption that β_1 and β_2 are of equal value but opposite in sign:

$$Y = \alpha + \beta\ \mathrm{RIE} + \varepsilon \qquad (5)$$

$$Y = \alpha + \beta\, E_{\mathrm{SMOKING}} - \beta\, E_{\mathrm{SWEETS}} + \varepsilon \qquad (6)$$

If implicit sweets attitudes do not predict the criterion, then the estimate of β in Eqs. (5) and (6) typically will be an intermediate value between 0 (the true value of β_2) and the true value of β_1. The standard technique of regressing criteria onto IAT scores will therefore not represent either β_1 or β_2, so the strategy of using a "control category" is problematic. Imagine, for instance, that implicit evaluations of smoking exert a strong influence on smoking behavior but that implicit evaluations of sweets have no effect at all on this same criterion. If one regresses smoking behavior onto the smoking–sweets IAT, then one might incorrectly conclude that implicit evaluations are a moderate or weak predictor of smoking behavior, even though they are a strong predictor (see Swanson et al., 2001).

There is a scenario, however, where a control category may be useful and this logic was incorporated into the SAT. If one assumes that the evaluation of the control category is constant for all study participants, then the IAT may be viable because variations in the relative implicit evaluation will actually reflect a single evaluation, not the difference between two evaluations. That is, if the relative implicit evaluation can be represented as

$$\mathrm{RIE} = E_{\mathrm{SMOKING}} - E_{\mathrm{SWEETS}}$$

where E_{SWEETS} is a constant, k, then we obtain

$$Y = \alpha + \beta\ (\mathrm{E}_{\mathrm{SMOKING}} - k) + \varepsilon$$

and the value of β is identical to that of the equation

$$Y = \alpha + \beta\ E_{\mathrm{SMOKING}} + \varepsilon$$

because subtracting a constant from a predictor does not alter its regression coefficient. We think it unlikely that evaluations of the concept of sweets function as a constant, so Study 2 sought a category that might have this quality for use in the SAT.[13]

Study 2

Respondents

The total sample consisted of 170 White undergraduate psychology students who participated for course credit. Following IAT conventions, 20 participants were eliminated for having greater than 15% error rates. One participant was eliminated from the final analyses due to a failure to complete the explicit measures. The final sample thus consisted of 149 participants (86 female, 63 male), ages 17–24 (mean = 18.97, SD = 1.36).

Measures

Simple association test

The SAT presented stimuli from six different categories on a computer screen. Two of the categories were the attitude objects used in the race IAT: *Blacks* (African American names) and *Whites* (European American names). Two of the categories were the evaluative anchors commonly used in the race IAT: *Positive* (positively valenced words) and *Negative* (negatively valenced words). In addition, the test used two control categories that were hypothesized to have evaluations that were near neutral and constant. The two control categories used were *furniture* (e.g., "table" and "desk") and *middle* (e.g., "midpoint" and "halfway").

To isolate the four association strengths, the SAT had four separate association tasks. In each task, only one

[13] Note that the crucial assumption is that of constancy, not neutrality.

attitude object (e.g., African Americans) and one evaluative category (e.g., positive) was used to create one joint category, which was then contrasted with the joint category created by pairing the two control categories. As an example, one block of tasks required participants to categorize stimuli via a key press as being either "African American or Positive" or as "Furniture or Middle." The speed of responding on this task was used to represent the strength of the association for "Blacks and Positive".

Four association strengths were assessed: *Blacks with positive* (B+), *Blacks with negative* (B−), *Whites with positive* (W+), and *Whites with negative* (W−). In addition to these experimental blocks, the task involved a set of practice blocks. The order in which these blocks were completed was based on protocols common in the IAT. The first block was a practice block consisting of 20 trials distinguishing evaluative anchors (positively and negatively valenced words). The second block was a practice block consisting of 20 trials distinguishing attitude objects (African American and European American names). The third block was a final practice block consisting of 20 trials distinguishing control categories (furniture and middle stimuli). The next four experimental blocks each involved 40 trials distinguishing joint categories. The order in which participants completed the different types of joint categories was randomly determined.

Following convention in the IAT literature, response latencies for the experimental trials were Winsorized to range from 300 to 3000 ms. We then applied a log-transformation to the reaction times. In addition, those participants who had error rates on more than 15% of the trials were eliminated. For ease of interpretation, the transformed scores on experimental trials were reverse coded so that higher scores indicated stronger association strengths. As in Study 1, we created multiple indicators of each association strength to allow for structural equation modeling. The first 3 responses in each experimental trial was eliminated to reduce noise during early responding. This left 36 transformed response latencies that could represent each of the four association strengths. With each of these trials, four separate indicators were created by randomly dividing the 36 trials into four subsets of nine trials and then calculating the average (transformed) response latency for each of the four groupings.

General processing speed

A measure of general processing speed also was obtained, but we used a somewhat different strategy than in Study 1. We had respondents complete a standard IAT task where the categories were flowers and insects and the attribute dimensions were positive and negative (Greenwald et al., 1998). Reaction time scores were randomly selected from these trials and grouped into four indicators. The mean of each indicator was used to model the influence of the latent construct, general processing speed.[14] We adopted this strategy because the joint categories in the IAT resemble those in the SAT and so they might provide a more valid estimate of the processing demands in this test (see McFarland & Crouch, 2002). As with the SAT, we Winsorized, log transformed and eliminated respondents with greater than 15% error rates.

Criterion measure

We also obtained a measure of *symbolic racism* (Henry & Sears, 2000). This was used as a criterion for prediction in some of our models. Symbolic racism is a form of prejudice that is thought to emanate from abstract, moralistic reasoning. Items attribute racial disparities to differences in work ethic and assess concerns for distributive justice. Based on the theory surrounding this measure, it seemed reasonable that it might be influenced by implicit evaluations related to the race IAT (Kinder & Sears, 1981; McConahay & Hough, 1976).[15] Using this score as a criterion helped us highlight the greater range of causal models that can be tested with the new measure.

Results

SAT measurement model

We first tested a number of theoretically meaningful models to evaluate the SAT factor structure. The first model tested the assumption that implicit evaluations of Blacks and Whites could be represented as separate bipolar dimensions. The response latencies for B+ and B− were thus assumed to have a single latent variable underlying them and the response latencies for W+ and W− were assumed to have a different latent variable underlying them. General processing speed was included in the model, in a manner consistent with Fig. 5. The pathways from the latent variable of general processing speed to the 8 indicators of implicit evaluations toward Blacks were constrained to be equal to one another, as were the comparable pathways to the 8 indicators of implicit evaluations toward Whites. The variances of the latent variables were fixed at 1.0. Evidence of good fit for this model justifies the view that implicit evaluations of racial groups are bipolar in form. The model yielded an unsatisfactory fit to the data (χ^2 (165) = 416.137, $p < .05$; GFI = 0.75; CFI = 0.85; RMSEA = 0.10; p value for close fit < .05; standardized RMR = 0.08). More liberal models that varied constraints on general processing speed also yielded poor model fit (for reasons that will become more apparent shortly). This argues against the view that the positive and negative

[14] It is worth noting that analysis of the bugs/flower IAT revealed the same factor structures as was found in the math/arts IAT in Study 1.

[15] Our goal here is not to make a strong case that implicit racism does cause symbolic racism. Our goal is simply to illustrate the methodological issues involved if one were to try to test theoretically meaningful criterion models using implicit measures.

implicit associations are simply bipolar opposites for a given attitude object.

The second model we tested distinguished between the four latent variables, B+, B−, W+ and W−. Each of these latent variables was estimated by four separate indicators. We covaried general processing speed out of each construct and examined the correlations between the residuals, using the modeling strategy in Fig. 6. This model yielded good fit to the data (χ^2 (160) = 164.66, $p < .38$; GFI = 0.91; CFI = 0.99; RMSEA < 0.02; p value for close fit > .05; standardized RMR = 0.035).

Correlation matrix

From this model, we analyzed the correlations between the four implicit attitude constructs (partialling out general processing speed) to gain insight into the implicit attitude structure. The correlation between B+ and B− was small, $r = 0.03$ (ns), as was the correlation between W+ and W−, $r = 0.08$ (ns). These values reinforce the view that implicit evaluations within racial groups are not bipolar in nature. Rather, the positive and negative implicit evaluations of the two different racial groups appear to be independent of one another, as is sometimes the case with explicit attitudes (Cacioppo & Berntson, 1994; Cacioppo, Gardner, & Berntson, 1997). This lack of an association also helps explain why the two IAT component judgments did not load onto a single implicit attitude factor in Study 1.

Results also revealed a small and significant positive correlation between B+ and W+, $r = 0.22$, and a significant positive correlation between B− and W−, $r = 0.32$ ($p < .01$). This result suggest that people who associate Blacks with negative characteristics also tend to associate Whites with negative characteristics and that people who associate Blacks with positive characteristics also tend to associate Whites with positive characteristics. This finding is hard to reconcile with the notion that implicit racial categories form "complementary pairs," as suggested by Greenwald & Farnham (2000). Rather, it suggests that some respondents have generally positive or generally negative associations of different types of people in the world, a bias that has been observed in the literature on explicit evaluations (Anderson, 1981).

The only remaining significant correlation was between W− and B+, $r = .37$ ($p < .01$). The direction of this correlation suggests that it is somewhat meaningful to construct an "incompatible judgment" for the race IAT (because this task assesses the degree to which participants associate Whites with positive and Blacks with negative). But the task also links W+ and B−, which were negligibly correlated.

Criterion prediction

Because analyses revealed that the SAT tapped four distinct associations, our causal modeling tested the independent influences of these four implicit race attitudes on symbolic racism. We did this by conducting a latent variable multiple regression analysis in which we regressed symbolic racism scores onto the four latent variables in the SAT that represent implicit associations and the latent variable representing general processing speed. Symbolic racism was modeled with a single indicator. The overall model fit was satisfactory (χ^2 (175) = 177.03, $p < .44$; GFI = 0.91; CFI = 0.99; RMSEA < 0.02; p value for close fit > .05; standardized RMR = 0.04). The multiple correlation was 0.31. The standardized regression coefficient for B− was the only significant predictor of symbolic racism ($B = 0.23$, $p < .03$). The nature of this effect was such that, the quicker participants were to associate Blacks with negative, the higher their symbolic racism scores. The standardized regression coefficients for the other implicit evaluations all were non-significant (Bs ranging from −.01 to .07, ps > .20), as was the coefficient for general processing speed, $B = 0.07$, $p < .40$. This pattern of results suggests that the tendency to implicitly associate Blacks with negative is the primary predictor of symbolic racism.

Discussion

Study 2 presented a measure that was designed to address some of the limitations of the IAT that were highlighted in Study 1. Our "simple association test" conceptually distinguished between the four key associations that IAT theorists emphasize in the race IAT. Rather than imposing an arbitrary aggregation of these associations and then differencing, we maintained each association as a separate entity that could contribute to criterion prediction in its own right. We elaborated a strategy for obtaining multiple indicators of each latent association and then we formally tested a measurement model that affirmed the viability of the indicators as well as the discriminant validity of the associations after controlling for general processing speed. The value of not combining the four associations into a single score can be seen in their differential impact on the criterion. We found that only negative associations with Blacks had a significant regression coefficient for predicting ratings on the symbolic racism scale. The traditional race IAT would obscure this dynamic.

General discussion

Decoding the IAT

The IAT has generated considerable theoretical and methodological interest in recent years. Despite the enthusiasm for the approach, a closer analysis reveals that the IAT framework requires restrictive causal assumptions and makes measurement assumptions that can be questionable. Because the IAT is focused on relative implicit evaluations at the conceptual level, theorists who use the IAT to predict psychological criteria must assume that the implicit evaluations of two distinct attitude objects combine in an additive fashion to impact the criterion of interest. They also must assume that the two implicit evaluations exert influences on the criterion that are equal in strength but opposite in sign. Study 1 revealed a case in which such assumptions appear not to be warranted. Study 2 showed the value of developing new measures that do not impose these restrictions. We found support for a causal model that views symbolic racism as a set of beliefs that is influenced primarily by the tendency to associate Blacks with negative characteristics, not with the tendency to associate Blacks with positive characteristics or with the tendencies to associate Whites with either positive or negative characteristics. This effect would not be uncovered with standard IAT measures.

In terms of measurement, IAT researchers have not made their measurement models explicit. We derived and tested several a priori plausible measurement models after "decoding" the IAT item format (Fig. 1). One of these models did not take into account general processing speed but others did. Study 1 found that none of the models yielded a satisfactory representation of IAT data. The only measurement model that did fit the data had a two-factor structure that arbitrarily grouped two of the four IAT associations into one cluster (e.g., M+ and A−) and the other two associations into a second cluster (e.g., M− and A+). Contrary to the traditional IAT practice of differencing these components, we found evidence that the components were not equally influential on our criterion. Study 2 suggested an approach for developing new implicit measures and provided some clues as to why the two IAT component scores failed to load on a single relative attitude dimension. We think this new approach offers promise, but we admit work on this measure has just begun.[16]

Limitations

Like any study, the present research has limitations that must be kept in mind when evaluating its implications. One limitation is that we based our analyses on psychometric models we derived as being a priori plausible, rather than being based on a psychometric model advocated by IAT researchers. This is because IAT researchers have not sufficiently elaborated such a model. This problem extends beyond psychometrics and also becomes relevant at the conceptual level. More specifically, the meaning of a "relative implicit evaluation" is ambiguous. One can use a number of different mathematical functions to represent a construct that is "relative." In Eq. (1), we used a difference function for the two implicit evaluations, such that $\text{RIE} = E_W/E_B$. But, perhaps another function better captures the meaning of relative. One alternative mathematical representation would be a ratio function, where $\text{RIE} = E_W/E_B$. In contrast to a difference function, this function implies an interactive causal model.[17] We did not examine the plausibility of this causal model in Study 1 because we found little evidence that IAT researchers meant to imply a ratio function, whereas they have made statements that imply a difference function (see Footnote 1).

Studies 1 and 2 presented new ways of dealing with general processing speed, but the need for such methods points to a limitation inherent in all response latency measures. The traditional method for "factoring out" processing speed relied on simple differencing (Greenwald et al., 1998). We showed how this approach can only remove processing speed confounds under restrictive conditions (see Appendix A). More recent approaches have relied on somewhat arbitrary scoring algorithms. A better approach is to model the influence of general processing speed directly in the data and to test the influence of implicit evaluations on criteria after this influence has been controlled. This was illustrated in Studies 1 and 2. It is important to note, however, that general processing speed remains a challenge for the IAT, the SAT and all response latency measures of attitudes. No analytic method will change the fact that the largest source of variation in the component parts of these measures probably is systematic error and that variation due to processing speed sometimes may reflect meaningful variation rather than simple method variance.

Future research on implicit attitudes

It might be argued that our concerns regarding measurement and causal models are allayed by evidence that the IAT and other implicit measures often are predictive of meaningful criteria (e.g., Asendorpf, Banse, & Muecke, 2002 Florack, Scarabis, & Bless,2001; Florack et al., 2001; Hugenberg & Bodenhausen, 2003; McConnell & Leibold, 2001; Nosek et al., 2002; Poehlman, Uhlmann, Greenwald, & Banaji, 2000; Poehlman et al., 2004). However, the issue at hand is not whether the IAT can predict criteria. The issue at hand is whether the results of criterion-prediction studies are subject to meaningful interpretation and capture the true underlying causal dynamics. To wit, Study 1 investigated a strong IAT–criterion relation-

---

[16] Numerous threats to validity may apply to the SAT. It is possible that our assumption of a constant "control category" is wrong. It may be, for example, that the "control category" is framed as negative when it is contrasted with positive and as positive when it is contrasted with negative (see Karpinski, 2004). This needs to be explored. We also think more research is needed to show that association strength can meet formal definitions of an "attitude" (see Karpinski & Hilton, 2001; Fazio & Olson, 2003; Fiedler et al., 2004; Rothermund & Wentura, 2001). Although the results of Study 2 are suggestive, our goal is not to assert that the SAT is a valid measure of implicit attitudes. We think it is premature to say that about any implicit association measure.

[17] For a ratio function, $\text{RIE} = E_W/E_B$ (where E_W is the implicit evaluation of Whites and E_B is the explicit evaluation of Blacks). To predict a criterion with this function, the simple linear model, $Y = \alpha + \beta\,\text{RIE} + \varepsilon$, or, $Y = \alpha + \beta\ (E_W/E_B) + \varepsilon$. This is the same as, $Y = \alpha + \beta\,(E_W)(1/E_B) + \varepsilon$. The presence of the product term implies a bilinear interaction between E_W and E_B, which is quite distinct from the additive function that characterizes differencing.

ship, but it also suggested that the types of interpretations that researchers have imposed in the past may not be warranted.

We think implicit measures of attitudes represent an interesting throw-back to other research paradigms from the past 50 years that were aimed at deriving indirect measures of attitudes that are not under the volitional control of the individual. These include unobtrusive observation (Webb, Campbell, Schwartz, & Sechrest, 1966), the bogus pipeline (e.g., Jones & Harold, 1971), galvanic skin response (Rankin & Campbell, 1955), pupillary response (Hess & Polt, 1960), and facial electromyographic activity (Cacioppo & Petty, 1979). Many of these efforts showed initial promise but ultimately suffered from one or more non-trivial drawbacks that undermined their widespread adoption. Our hope is that researchers will consider these examples and then take a step back to design new implicit measures that are conceptually flexible and that have explicit, well developed, and testable (i.e., falsifiable) psychometric models that can accommodate the complex dynamics of social behavior. With such measures in hand, it will make sense to test and interpret the relationship between implicit attitudes and psychological criteria.

Appendix A. Controlling general processing speed

This appendix discusses two strategies for removing the effects of general processing speed. The first is the "traditional" strategy of computing a simple difference between the compatible and incompatible response times (Greenwald et al., 1998). The second is the new scoring algorithms suggested by Greenwald et al. (**2003, Nosek, Greenwald and Banaji, 2005). Both methods attempt to expunge processing speed from the IAT score, rather than statistically controlling for it in the context of a broader model being tested with a given data set.

A.1. Conditions where simple differencing factors out general processing speed

A necessary condition for difference scoring to "control for" general processing speed successfully is that the compatible and the incompatible judgments must be influenced to the same degree by it. This can be shown with simple algebra. The path model in Fig. 3 implies the following measurement equations:

$$\mathrm{CRT} = \alpha_2 + \lambda_1\ \mathrm{RIE} + \lambda_3\ \mathrm{GPS} + \varepsilon_2,$$

$$\mathrm{IRT} = \alpha_1 + \lambda_2\mathrm{RIE} + \lambda_4\ \mathrm{GPS} + \varepsilon_1.$$

For ease of presentation, we assume that the relative implicit evaluation (RIE) has an equal but opposite influence on the compatible reaction time (CRT) and incomptaible reaction time (IRT), and so $\lambda_2 = \lambda_1$. (Results of Study 1 argue against this simplifying assumption, but the conclusions drawn in this appendix do not change if this assumption is violated.) The key assumption in our analysis is that general processing speed (GPS) exerts an equal influence on CRT and IRT, and so $\lambda_3 = \lambda_4$. Under this set of conditions, the IAT difference score becomes

$$\begin{aligned}\mathrm{IRT} - \mathrm{CRT} &= [\alpha_1 + \lambda_1\ \mathrm{RIE} + \lambda_3\ \mathrm{GPS} + \varepsilon_1] - [\alpha_2 \\ &\quad + -\lambda_1\ \mathrm{RIE} + \lambda_3\ \mathrm{GPS} + \varepsilon_2] \\ &= (\alpha_1 - \alpha_2) + 2\lambda_1\ \mathrm{RIE} + (\varepsilon_1 - \varepsilon_2).\end{aligned}$$

The only factors creating variation in the difference score in the last equation are the relative implicit evaluation and random measurement error. Thus, the difference score will yield an estimate of an individual's relative implicit evaluation that is free of general processing speed, provided that general processing speed exerts equal influence on the two component scores.

To date, the assumption that general processing speed has an equal influence on CRT and IRT has not been tested, but it seems likely that it often is not justified. One source of doubt derives from research on task performance and task difficulty. A large literature suggests that general abilities (e.g., general processing speed) influence performance on specific tasks (e.g., reaction time for IAT tasks) differentially as a function of task difficulty (Ackerman, 1986, 1987). For researchers to argue that general processing speed exerts the same influence on the compatible and incompatible tasks, they also must argue that the two judgments are equal in difficulty. This seems to contradict the theory surrounding the IAT, which supposes that incompatible categorizations typically will be harder for people to make than compatible categorizations (Greenwald et al., 1998).

Results of recent studies suggest that differences in task difficulty are, in fact, creating a processing speed confound in IAT scores. These studies indicate that overall reaction times correlate with IAT difference scores (Greenwald et al., 2003; Hummert, Garstka, O'Brien, Greenwald, & Mellott, 2002; McFarland & Crouch, 2003) and that conceptually unrelated IAT share method variance (McFarland & Crouch, 2003). This suggests that general processing speed may not be factored out by computing a difference score.

A.2. New scoring algorithms

An alternative strategy for factoring out general processing speed has been suggested by Greenwald et al. (2003). These authors introduced a scoring algorithm that they feel effectively does so and that also yields a better index of the relative implicit evaluation. More specifically, they evaluated several different scoring methods and favored those methods that minimized the correlation between an index of general processing speed and the IAT difference score and that maximized the correlation between the IAT and explicit measures of relative attitudes.

One problem with the approach of Greenwald et al. (2003) is that researchers do not know the extent to which the new scoring procedure truly removes the influence of general processing speed for any given research application. The observed correlations of the newly scored IAT and general processing speed that Greenwald et al. reported were low for the particular topics that they studied. However, a researcher will not know if this is true for his or her particular IAT measure unless an index of general processing speed is obtained and its correlation with the IAT score examined. Once a researcher has gone to the effort of measuring general processing speed to answer this question, it is a dubious strategy to use the somewhat *ad hoc* and inexact methods for controlling processing speed advocated by Greenwald, Nosek and Banaji. The better approach is to obtain multiple indicators of general processing speed (so as to adjust for random measurement error) and then use these in a structural equation model to explicitly partial out its effects in a formal system of theoretically guided linear equations.

Another issue with the scoring method is the algebraic strategy it uses to control for general processing speed. The primary way that this is accomplished is by dividing the IAT difference score for a respondent by the standard deviation of that person's response latencies computed across compatible and incompatible tasks (see Greenwald et al., 2003, for details). This takes the general form of (IRT–CRT)/σ. If dividing by the standard deviation controls for general processing speed, then σ should be highly correlated with an independent measure of it. In our data, the correlation of σ with indicators of general processing speed was approximately 0.45, suggesting that σ may reflect factors other than general processing speed. The result will be inexact control of true general processing speed. One also should provide a theoretical account for why it is preferable to divide IAT scores by an indirect estimate of processing speed, σ, rather than statistically controlling for a more direct estimate of it in the context of a broader model.

Finally, we note that some of the criteria used by Greenwald et al. to settle on a scoring method are controversial. For example, Greenwald et al. suggested that an optimal IAT scoring procedure should maximize the correlation between the measures of implicit evaluations and measures of explicit evaluations. This represents one plausible view of implicit attitudes—that they reflect evaluations that are conceptually related to their explicit counterparts. It is not the only plausible view. For example, Banaji (2001) argued that implicit and explicit measures should be independent of one another because in "the research world in which conscious and unconscious attitudes are seen to be conceptually distinct, the low correlation or lower correlation between measures within the same family is taken as evidence of validity, not a challenge to it." (pp. 143).

A.3. Method variance or meaningful variance?

An additional complication of removing the influence of general processing speed is that general processing speed can have a meaningful relationship rather than an artifactual relationship to the theoretical constructs one wishes to represent. For instance, suppose that general processing speed is reflective of intelligence, more or less. Suppose also that intelligence is related to implicit self-esteem such that people with higher intelligence have higher implicit self-esteem. In this case, one expects general processing speed to be correlated with the *latent* construct of interest. If a scoring algorithm removes the influence of general processing speed on the *observed* measure, it also would remove some of the true variation in implicit self-esteem, namely that portion of implicit esteem that covaries with intelligence. The result can be a distorted self-esteem index that will bias parameter estimates in the broader model under investigation. Scoring corrections that try to expunge general processing speed from the observed IAT score force the theorist to assume that the true implicit attitude does not covary at all with general processing speed or constructs that are correlated with general processing speed. This is problematic because there is empirical literature linking constructs such as intelligence, academic achievement, and other cognitive abilities to many of the types of variables studied in IAT research. These include self-esteem (Greenwald & Farnham, 2000), racist attitudes (Monteith, Voils, & Ashburn, 2001; Ottaway, Hayden, & Oakes, 2001; Rudman, Greenwald, Mellott, & Schwartz, 1999), homophobia (Banse, Seise, & Zerbes, 2001), sexual risk taking (Marsh, Johnson, & Scott-Sheldon, 2001), and racial discrimination (McConnell & Leibold, 2001). These literatures suggest that one might expect correlations between general processing speed and valid IAT measures of these constructs and that these correlations would not be artifactual but conceptually meaningful.

A.4. Conclusion

Given the above, we believe that future research on implicit evaluations will be served best if researchers do not attempt to control for processing speed confounds either by differencing or by employing the new scoring algorithm in Greenwald et al. (2003). Instead, researchers should pursue the general strategies we introduce here of measuring general processing speed and then modeling its effects in the data as dictated by the more general role of general processing speed in the broader model being tested.

References

Ackerman, P. L. (1986). Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. *Intelligence, 10*(2), 101–139.

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*(1), 3–27.

Anderson, N. (1981). *Methods of information integration*. New York: Academic Press.

Asendorpf, J. B., Banse, R., & Muecke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83*(2), 380–393.

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, J. S. Nairne, I. Neath, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift fuer Experimentelle Psychologie, 48*(2), 145–160.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62*, 893–912.

Blanton, H., & Jaccard, J. (in press). Arbitrary metrics in psychology. *American Psychologist*.

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*(5), 760–773.

Cacioppo, J. T., & Petty, R. E. (1979). Attitudes and cognitive response: An electrophysiological approach. *Journal of Personality and Social Psychology, 37*, 2181–2199.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review, 1*, 3–25.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115*(3), 401–423.

Carver, C. S. (1989). How should multifaceted personality constructs be tested. Issues illustrated by self-monitoring, attributional style, and hardiness. *Journal of Personality and Social Psychology, 56*(4), 577–585.

Christie, C., Blanton, H., & Jaccard, J. (2005). Factor structure of measures of implicit association. Unpublished Manuscript

Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin, 74*(1), 68–80.

De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology, 50*, 77–85.

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

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62–68.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. New York: Harcourt and Brace.

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

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of altidues. *Journal of Personality and Social Psychology, 50*(2), 229–238.

Fiedler, K., Messner, C., & Matthias, B. (2004). Unresolved problems with the "I", the "A" and the "T": Logical and psychometric critique of the implicit association test (IAT). Unpublished manuscript.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading Mass: Addison-Wesley.

Florack, A., Scarabis, M., & Bless, H. (2001). When do associations matter. The use of automatic associations toward ethnic groups in person judgments. *Journal of Experimental Social Psychology, 37*(6), 518–524.

Gonzales, P. M. (2002). The use of difference scores in the implicit association test. Unpublished doctoral dissertation, The State University of New York at Albany.

Greenwald, A. G, & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*(6), 1022–1038.

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*(6), 1064–1480.

Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie, 48*, 85–93.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Scoring procedures to improve implicit association test measures. *Journal of Personality and Social Psychology, 85*, 197–216.

Griffin, D., Murray, S., & Gonzalez, R. (1999). Difference score correlations in relationship research: A conceptual primer. *Personal Relationships, 6*(4), 505–518.

Henry, P. J., & Sears, D. O. (2000). The symbolic racism 2000 scale. *Political Psychology, 23*(2), 253–283.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*, 349–350.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science, 14*(6), 640–643.

Hummert, M. L., Garstka, T. A., O'Brien, L. T., Greenwald, A. G., & Mellott, D. S. (2002). Using the Implicit Association Test to measure age differences in implicit social cognitions. *Psychology and Aging, 17*, 482–495.

Jaccard, J., & Wan, C. (1996). *LISREL approaches to interaction effects in multiple regression*. Sage: Newbury Park.

Johns, G. (1981). Difference score measure of organizational behavior variables: A critique. *Organizational Behavior and Human Decision Processes, 27*(3), 443–463.

Jones, E. E., & Harold, S. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76*, 349–364.

Karpinski, A., & Steinman, R. B. (in press). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology Bulletin*.

Karpinski, A. (2004). Measuring self-esteem using the implicit association test: The role of the other. *Personality and Social Psychological Bulletin, 30*(1), 22–34.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the implicit association test. *Journal of Personality and Social Psychology, 81*(5), 774–788.

Kinder, D. R., & Sears, D. O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology, 40*(3), 414–431.

Klauer, K. C., & Mierke, J. (2004). Task-set inertia, attitude accessibility, and compatibility-order effects: New evidence for a task-set switching account for the IAT effect. *Personality and Social Psychology Bulletin*, in press.

Marsh, K. L., Johnson, B. T., & Scott-Sheldon, L. A. (2001). Heart versus reason in condom use: Implicit versus explicit attitudinal predictors of sexual behavior. *Zeitschrift fuer Experimentelle Psychologie, 48*(2), 161–175.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37*(5), 435–442.

McConahay, J. B., & Hough, J. C. (1976). Symbolic racism. *Journal of Social Issues, 32*(2), 23–45.

McFarland, S. G., & Crouch, Z. (2003). A cognitive skill confound on the Implicit Association Test. *Social Cognition, 20*, 483–506.

McFarland, S. G., & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition, 20*(6), 483–510.

Monteith, M. J., Voils, C. I., & Ashburn, N. L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition*, *19(4)*, 395– 417.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*(6), 625–666.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math not= me. *Journal of Personality and Social Psychology, 83*(1), 44–59.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*(2), 166–180.

Olson, M., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the implicit association test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*(5), 653–667.

Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (2001). Implicit attitudes and racism: Effects of word familiarity and frequency on the implicit association test. *Social Cognition, 19*(2), 97–144.

Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology, 48*, 609–647.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., et al. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12*(5), 729–738.

Poehlman, T. A., Uhlmann, E., Greenwald, A. G., & Banaji, M. R. (2004). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. Unpublished manuscript.

Rankin, R. E., & Campbell, D. T. (1955). Galvanic skin response to Negro and white experimenters. *Journal of Abnormal and Social Psychology, 51*, 30–33.

Rothermund, K., & Wentura, D. (2001). Figure-ground asymmetries in the Implicit Association Test (IAT). *Zeitschrift fuer Experimentelle Psychologie, 48*(2), 94–106.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test (IAT): Dissociating salience from associations. *Journal of Experimental Psychology: General, 133*, 139–165.

Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. K. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition, 17*(4), 437–465.

Steffens, M. C. (2004). Is the implicit association test immune to faking? *Experimental Psychology, 51*(3), 165–179.

Swanson, J. E., Rudman, L. A., & Greenwald, A. G. (2001). Using the implicit association test to investigate attitude–behavior consistency for stigmatized behavior. *Cognition and Emotion, 15*, 207–230.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally.