

## A COGNITIVE SKILL CONFOUND ON THE IMPLICIT ASSOCIATION TEST

Sam G. McFarland and Zachary Crouch

We hypothesized that scores on the Implicit Association Test (IAT) are confounded with a general cognitive ability of how quickly one can process information when the IAT categories seem incongruent compared to when they are congruent. Across four studies, two IATs on irrelevant dimensions (e.g., delicious–happy) were substantially correlated with IATs assessing prejudice and self-esteem, confirming the general skill confound: Those who lack this skill are biased toward higher prejudice and lower self-esteem IAT scores. However, IATs with just two exemplars in each category were much less affected by this confound than were those with more exemplars. The cognitive skill was associated specifically with the difficulty in responding when categories were incongruent and to the difficulty of mentally switching sets in the middle of an IAT.

In recent years, the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) has rapidly become a popular method for assessing attitudes that individuals might not recognize in themselves or might be unwilling to express. In addition to its initial use in measuring prejudice, the IAT has recently been used to measure implicit self-esteem (Greenwald & Farnham, 2000), dysfunctional beliefs (de Jong, Pasman, Kindt, & van den Hout, M. A., 2001), fears of snakes and spiders (Teachman, Gregg, & Woody, 2001), depression (Gamar, Segal, Sagrati, & Kennedy, 2001), shyness (Asendorpf, Banse, & Mücke, 2002), and anxiety (Egloff & Schulke, 2002).

---

Correspondence about this article may be addressed to Sam McFarland at Department of Psychology, Western Kentucky University, Bowling Green, Kentucky 42101, or by e-mail to [sam.mcfarland@wku.edu](mailto:sam.mcfarland@wku.edu). We thank Arthur Kendall for his thoughtful insights that helped define this research project.

The promise of the IAT is that it offers a method for assessing prejudice and other attitudes that is immune to both self-denial and self-presentational concerns. But these immunities do not mean that the IAT is free from other method artifacts. In its general format, the IAT measures the degree to which response speeds are slower when categories are incongruent (to the individual) compared to when they are congruent. To the degree that this speed difference reflects a general cognitive inability to suppress or ignore incongruence, that inability might confound IAT scores. For example, those lacking this ability would be biased toward higher IAT prejudice scores. This is the general issue explored here.

The IAT is a response latency test taken on a desktop computer. Participants choose whether a stimulus presented in the middle of the screen belongs in the category listed on the left of the screen or that on the right by tapping the "a" or "5" (on the number pad) keys as rapidly as possible. Responses are timed in milliseconds. For example, to assess implicit anti-Black racial prejudice, participants must press one key for exemplars that are in the White category (such as a White face or stereotypical White name) and the other key for exemplars that are Black (a Black face or stereotypical Black name). The Black-White dimension is usually cross-matched with a positive (e.g., "joy") vs. negative ("agony") dimension, although moral ("kind") vs. immoral ("hateful") has also been used. For half of the test, White is paired with the positive category and Black is paired with the negative one. For the other half, Black is paired with positive and White with negative. An individual's implicit prejudice score is the difference in mean response time between these two pairings. Prejudice is indicated to the degree that responses to the [White-negative, Black-positive] pairings are slower than to the [White-positive, Black-negative] pairings. The IAT has been used to measure other implicit prejudices, including sexism, ageism, antihomosexuality, and antiobesity. Psychometrically, the IAT appears to have adequate internal consistency and temporal stability (Cunningham, Preacher, & Banaji, 2001), although these are often weak in comparison to self-report, explicit measures (such as Likert scales) of the same constructs. Also, several procedural artifacts have little effect upon IAT responses, including the words chosen to repre-

sent the categories, as long as they are somewhat familiar (Ottaway, Hayden, & Oakes, 2001), the number of exemplars in the categories, and intertrial interval (Greenwald et al., 1998).

Prejudices measured by the IAT usually correlate only in the .20s with explicit, Likert-type scales. McFarland and Mattern (2002) recently found that IAT measures of prejudices against Blacks, women, homosexual persons, foreigners, and the poor all intercorrelated substantially, yielding a single factor of "generalized implicit prejudice" that paralleled generalized explicit prejudice. However, generalized prejudice measured by the two methods correlated only .23.

However, some individuals may have greater difficulty than others in responding to incongruent categories relative to congruent ones, independent of the specific content of the IATs. The empirical question is whether IAT scores on unrelated constructs correlate substantially with IAT scores of implicit prejudice or other constructs of interest. To the degree that they do, IAT measures will be shown to be confounded with a general cognitive skill of how quickly one can correctly categorize exemplars presented in incongruent categories as compared to when they are presented in congruent ones.

To demonstrate the feasibility of the IAT as a measurement tool, Greenwald et al. (1998, Experiment 1) tested two nonsocial IATs. In the first, well-known flowers (e.g., "iris") and insects ("wasp") were matched and cross-matched with pleasant ("love") and unpleasant ("filth") words. In a second, musical instruments and weapons were used with the pleasant and unpleasant words. Participants responded faster to the [flowers-pleasant, insects-unpleasant] and [musical instruments-pleasant, weapons-unpleasant] pairings than to the opposite, incongruent pairings. Greenwald et al. reported that these two nonsocial IATs correlated .58 and suggested that this correlation may reflect a systematic method variance. But they did not explore the nature of this method variance nor investigate how much it contaminates IAT measures of prejudice or other constructs.

Should a general cognitive skill affect IAT responses, what is its likely nature? Jones and Jacoby's (2001) model of dual memory processes, used to explain false memories, may provide a useful framework for explaining IAT responses as well. In this

model, a fast and automatic process of familiarity is followed by a slower, controlled process of recollection. In false memory tasks, participants must retain features of previously studied compound words to correctly identify the words but must suppress the impulse to falsely identify words that contain these features (e.g. "heart" must be retained to correctly recall "heartburn," but the impulse to identify "heartbeat" as a previously studied word must be suppressed). The automatic process is sufficient to identify the previously studied word, but if a feature is presented in an unstudied word, false memories will occur if the automatic process is unchecked by the controlled process.

In parallel, in the congruent IAT condition, familiarity need not be checked by controlled processes in order for a correct identification to occur. "Tasty" and "optimism" both elicit automatic positive responses, and one may not even need to place them in the categories of "delicious" and "happiness" to respond correctly. But during incongruent presentations, several steps of controlled processing are clearly required. If anything, the controlled processing required by the IAT is more complex than processing necessary to suppress false memories. A participant must (a) suppress responding to the exemplars based on their simple positivity or negativity, (b) place them in their correct categories, (c) recall whether the categories are on the right or left, while (d) suppressing the incongruence of placement of the categories.

However, for two reasons it appears unlikely the IAT reflects *only* this general cognitive skill. First, IAT measures of prejudice generally correlate significantly, if weakly, with explicit measures of the same prejudices (e.g., Greenwald et al., 1998; McFarland & Mattern, 2002). More importantly, several recent studies have found that IAT-assessed attitudes predict attitude-relevant behaviors. McConnell and Leibold (2001) found that IAT-assessed racial prejudice led to more negative nonverbal interactions (e.g., less speaking and smiling, more speech hesitation and errors, more negative ratings of the interaction) with Black as compared to White experimenters. Rudman and Glick (2001) found that IAT-assessed (but not explicitly expressed) gender stereotypes predicted discrimination in rating agentic (competitive and assertive) female job applicants. In non-prejudice domains,

Asendorph, et al. (2002) found that an IAT-assessed association of the self with shyness predicted shy body language (e.g., gaze aversion, tense body posture) in a get-acquainted session; Egloff and Schmulke (2002) showed that an IAT associating the self and anxiety predicted nervous body language (e.g., nervous hand and body movements, speech dysfluency) during a stressful speech. Nevertheless, IAT scores may still be distorted substantially by individual differences in the cognitive skill described above. To date, no tests of this potential confound have been reported.

The hypothesized confound, should it exist, could be due to either responding faster when categories are congruent or slower when they are incongruent. Given the unique cognitive requirements of the IAT outlined above, it seemed more likely that the latter would be the case, that the main individual difference would be that some participants can disattend quickly to category incongruence while others cannot. For that reason, it is expected that the hypothesized confound will primarily affect the speed of responding when categories are incongruent. These are the issues examined in this research.

## STUDY 1

### METHOD

The effects of the cognitive confound upon prejudice IATs were examined in Study 1.

*Materials.* Three IAT prejudice measures were used, including prejudice against Blacks, homosexual persons, and "foreigners," all taken from McFarland and Mattern (2002). The anti-Black measure, which had been adapted from Greenwald et al. (1998), consisted of eight stereotypical White female names (e.g., "Nancy") and eight stereotypical Black female names ("Latonya"). The antforeign measure contained ten names and words that one would associate with American (e.g., "Washington," "Buffalo ") and Foreign ("Gandhi," "Kangaroo"). The homosexual – heterosexual measure was merely two synonyms (e.g., gay and straight) and related exemplars for each word. The list was limited to only two words in order to avoid words with strong derogatory connotations (e.g., "fag"). For the three preju-

dice IATs, the matched evaluative lists were comprised of eight exemplars for moral (e.g., "honest") and for immoral ("cruel"). These categories were used instead of positive – negative because exemplars in the moral – immoral categories are more directly evaluative in the sense of good versus bad.

Two control IATs were used. One was developed that used the dimensions of delicious (e.g., "tasty") – not delicious ("rancid") and happiness ("optimism") – unhappiness ("hopeless"). While these dimensions are logically distinct, the pairings of delicious and happiness (both positive qualities) and not delicious and unhappiness (both negative) seem psychologically congruent, while the opposite pairings seem incongruent. For a second control IAT, Greenwald et al.'s (1998) flowers – insects measure was adopted and abbreviated. The exemplars used on all IATs are reported in the Appendix.

If the general cognitive skill described above is found to influence IAT measures, it will be useful to know if it correlates with Likert measures as well: Perhaps those who are slower at responding to incongruent pairings as compared to congruent ones are, in fact, more prejudiced. For that reason, Likert scales were used to measure the same three prejudices. Anti-Black racism was measured by eight items from McConahay's (1986) Old-Fashioned and Modern Racism Scales; these eight items were the four pro-trait and four con-trait items that loaded highest on the single factor of the items from these two scales. For antihomosexual attitudes, the four highest loading pro-trait and con-trait items on Kite and Deaux's (1986) Attitudes Toward Homosexuals Scale were used. For antforeign attitudes, a 12-item balanced ethnocentrism scale was adapted from Altemeyer's Manitoba Prejudice Scale (Altemeyer, 1988, p. 110); this scale measured a general rejection of a number of non-American outgroups, including Russians, Indians, Japanese, Arabs, and Asians.

*Sample.* The five IAT measures and the three Likert scales were administered to 81 undergraduate psychology students who received course credit for their participation. All were traditional American 18-to-20 year-old students. Because attitudes toward Blacks was one of the prejudices, the five Black students were dropped from the analyses. However, analyses repeated

including these five participants yielded virtually identical results.

*Procedure.* The students were tested individually. After reading and signing a consent form, the students completed the questionnaire and IAT tests. The order of the two was counterbalanced so that half completed each one first. The five IAT tests were presented in a constant sequence: flowers–pleasant, anti–Black prejudice, delicious–happiness, antihomosexual prejudice, and antforeign prejudice. Within each test, students received in order

- (a) 20 practice trials on each separate dimension (e.g., White–Black and moral–immoral),
- (b) 20 practice trials on the stereotype–congruent paired dimensions (e.g., White–moral vs. Black–immoral),
- (c) 40 test trials on the congruent paired dimensions,
- (d) 20 practice trials on the moral – immoral dimension with the terms switched to the opposite side of the screen,
- (e) 20 practice trials on stereotype–incongruent paired dimensions (e.g. Black–moral vs. White–immoral), and
- (f) 40 test trials on the incongruent paired dimensions.

Following Greenwald et al. (1998), the first two test trials (blocks c and f) were discarded, leaving 38 trials. The order of the IAT tests and of the congruent–incongruent order within each test were not varied. Previous studies have found that IAT effect sizes are generally larger when congruent pairs are presented first (Greenwald & Nosek, 2001), but order has been shown not to affect an IAT's correlations with other measures (McConnell & Leibold, 2001).

## RESULTS

Key punch error rates on the five IATs were highly negatively skewed and highly correlated across the five tests, median = .52,  $p < .001$ . Several participants had error rates that indicated extreme carelessness, if not downright randomness. An arbitrary—but logical—decision was made to delete participants whose error rates were greater than 33% averaged across the IATs. Using this stan-

TABLE 1. Mean Millisecond Latencies for the Five IAT Measures with "Matched" and "Mismatched" Evaluative Exemplars, Study 1

Category matches	Latency (milliseconds)	Cohen's <i>d</i>
Anti-Black Racism		
White-Moral and Black-Immoral	795 <sup>a</sup> , [936] <sup>a</sup>	1.31
White-Immoral and Black-Moral	1113, [996]	[.42]
Antihomosexuality		
Heterosexual-Moral and Homosexual-Immoral	805 <sup>a</sup> , [911] <sup>a</sup>	1.41
Heterosexual-Immoral and Homosexual-Moral	1025, [991]	[.51]
Antiforeign		
American-Moral and Foreign-Immoral	802 <sup>a</sup> , [887] <sup>a</sup>	1.06
American-Immoral and Foreign-Moral	1005, [990]	[.68]
Taste		
Delicious-Happy and Not Delicious-Unhappy	699 <sup>a</sup>	1.81
Delicious-Unhappy and Not Delicious-Happy	1322	
Flowers-Insects		
Flowers-Pleasant and Insects-Unpleasant	714 <sup>a</sup>	1.33
Flowers-Unpleasant and Insects-Pleasant	1019	

<sup>a</sup>The two means differ,  $p < .001$ , on one-sample  $t$ -test. The bracketed numbers are ANCOVA adjusted means and Cohen's  $d$ 's, controlling for the control IATs.

dard, 14 participants (18%) were deleted, leaving a final sample of 43 females and 19 males.<sup>1</sup>

*IAT-Assessed Measures.* As is commonly done (Greenwald et al., 1998), outlier trials with latencies longer than 3000 milliseconds or shorter than 300 milliseconds were recoded to these numbers. As shown in Table 1, the latency on each IAT was longer when the categories were incongruent, and each one had a significant  $t$ -test for single sample  $t$ -tests and a substantial Cohen's  $d$  (Cohen, 1988).

1. Greenwald and Farnham (2000) deleted participants with greater than 20% error rates, so the standard for keeping participants here was more lenient. Still, deleting 18% of the respondents for that reason may seem high, but certainly the validity of scores on a computer timed test are questionable when respondents punch the wrong keys a third of the time. However, results using the full White sample were essentially the same, except that the correlations among the five IAT tests were a bit lower: The median interIAT correlation on the reduced sample and full samples were .27 and .25, respectively. Using the total White sample would not have altered the essential results and conclusions of this report.



Replicating McFarland and Mattern (2002), the IAT measures of prejudice intercorrelated between .33 and .39,  $p < .01$  in all cases. Principal axis analyses of the six prejudice measures (three IATs and three Likert scales) yielded two factors, with each one comprised almost exclusively of the three measures associated with each method. Rotated obliquely, these two factors correlated .28.

Of interest for the present study, however, are the relations of the IAT prejudice measures with the two control IATs. The correlations between them, presented in Table 2, indicate that the responses to the two control IATs correlated significantly with IAT scores on two of the three measures of prejudice (one-tailed tests are reported because positive correlations were expected). Only scores on the antihomosexual IAT did not correlate with scores on either control IAT. However, the two control IATs correlated only .22,  $p < .05$ , with one another, .31 corrected for attenuation.

How great is this cognitive skill contamination of IAT measures of prejudice? Because of the relatively low reliabilities of the IAT measures for 38-item scales, and to estimate the degree of the cognitive confound for true scores, the significant correlations were corrected for attenuation (Nunnally, 1967, p. 219); these unattenuated correlations are presented in brackets in Table 2. The correlations were enhanced substantially. To estimate the true strength of this contamination, the standardized scores of the two control IATs were summed and this sum was correlated with each of the IAT prejudices. The magnitude of these correlations, presented in the right column of Table 2, shows that the effects of this cognitive ability upon responding to the IAT is substantial rather than inconsequential, accounting for 25% of the variance in IAT-assessed anti-Black and antiforeign prejudices. It should be noted, however, that correlations corrected for attenuation have large confidence intervals. Following procedures outlined by Forsyth and Feldt (1969), the 90% confidence interval for the unattenuated anti-Black correlation is .24 to .76; that for the corrected antiforeign correlation is .22 to .75.

It is also important to know how much of the large prejudice effect sizes reported in Table 1 are due to the cognitive artifact. Do the adjusted means for the incongruent and congruent pairings remain significantly different once the cognitive artifact is controlled? For each of the three prejudices, ANCOVA was used to calculate the adjusted means for the incongruent and congruent

TABLE 2. Correlations of the Prejudice and self-esteem IATs with the Control IATs

	Cognitive Confound IAT		
	Delicious (D) (.69, .81)	Flowers (F) (.75, .75)	Sum (D + F) (.71, .78)
Study 1			
Antihomosexual IAT (.73)	-.04	.12	.12
Anti-Black IAT (.77)	.29** [.39]	.28a [.37]	.37** [.50]
Antiforeign IAT (.65)	.25* [.37]	.26* [.37]	.33** [.49]
Study 2			
Antihomosexual IAT (.80, <i>n</i> = 89)	.40* [.50]	.08	.29* [.40]
Anti-Black IAT (.74; 8 exemplars; <i>n</i> = 37)	.30* [.42]	.55** [.73]	.54** [.72]
Anti-Black IAT (.85; 2 exemplars; <i>n</i> = 43)	-.07	.08	.10
Self-Esteem IAT (.87; 8 exemplars; <i>n</i> = 47)	.35* [.45]	.44** [.54]	.35** [.43]
Self-Esteem IAT (.81; 2 exemplars; <i>n</i> = 42)	.18	-.15	.03

*Note.* For Study 1, *n* = 62 for all comparisons. Numbers in parentheses are the alphas for the IATs. Those below Delicious, Flowers, and Sum are for Studies 1 and 2, respectively. Numbers in brackets are correlations corrected for attenuation. \*\**p* < .01; \**p* < .05.

pairings controlling for the control IATs and to examine whether these pairings remained significantly different. These adjusted means are presented in brackets in Table 1. In each case, the differences remain highly significant,  $p < .001$  in all cases, but the Cohen's-*d* on the adjusted means are much smaller than those on the unadjusted means.

The control IATs were not related to the Likert measures of prejudice. The only significant correlation (out of six) between a control IAT and the Likert measures (Flowers IAT with the Anti-Black scale) was  $-.22$ ,  $p < .05$ , suggested that those who had greater difficulty with the flowers-insects IAT were less prejudiced than others against Blacks, but this result may well be a chance anomaly.

It is appropriate to ask whether controlling for the flowers-insect and delicious-happy IATs affects the correlations among the prejudice IATs and their correlations with the explicit prejudices. The partial correlations between the three prejudice IATs were reduced, which now intercorrelated from  $.21$ ,  $p < .05$  to  $.29$ ,  $p < .01$ . The cognitive confound thus enlarged the correlations among the three prejudices, but McFarland and Mattern's (2002) finding of a generalized implicit prejudice appears to remain.

Controlling for the confounds had little effect upon the correlations between the IATs and their respective explicit measures.

The anti-Black IAT and prejudice scale correlation of .28,  $p < .05$ , became .30 and the antihomosexual IAT and scale correlation of .22,  $p < .06$  became .23 with the cognitive confounds controlled. The nonsignificant correlation between the antiforeign IAT and scale,  $r = .14$ ,  $p > .10$ , became just .16,  $p > .10$ , with the controls. The correlation between the implicit and explicit factor scores was changed from .28 to .29. These small changes were due to the fact that the control IATs were unrelated to the measures of explicit prejudice.

Finally, we examined whether the cognitive confound with IAT prejudice scores was due to slower responding when words were presented in incongruent categories, faster responding when presented in congruent categories, or both. The two control IATs were correlated with the separate response speeds to the congruent and incongruent presentations on the anti-Black and antiforeign IATs. The four correlations for congruent scores ranged from .05 to .16, *ns* in all cases. The comparable correlations for the incongruent scores ranged from .25 to .37,  $p < .05$  in all cases. Thus, the cognitive ability confound appeared to be specifically related to response speeds on the prejudices when words were presented in incongruent pairings.

## DISCUSSION

The results of this study showed that individual's scores on two of three IATs assessing prejudice were confounded by a general cognitive skill assessed by two control IATs on irrelevant, nonsocial dimensions. Further, the cognitive skill confound was specifically associated with response speeds on the incongruent pairings of the prejudice IATs.

While both control scales used in this study appear to reflect an ability to ignore incongruent categories, the low correlation between them shows that they are not identical. Perhaps the difference is simply their relative difficulty. Responding to the incongruent categories of (delicious – unhappy, not delicious – happy) was much more difficult than for the incongruent (flowers – unpleasant, insects – pleasant). The former do seem more incongruent than the latter, and they required an average of three-tenths of a second longer latency (see Table 1). Neverthe-

less, each one appears to tap the cognitive capacity that creates impurity in IAT-assessed prejudice.

Given the correlations of both control IATs with the other prejudices, their failure to correlate with the antihomosexuality IAT is puzzling. The significant correlations of the antihomosexuality IAT with the anti-Black IAT,  $r = .39$  and antiforeign IAT,  $r = .33$ , add to the puzzle. Only the small number of exemplars on the antihomosexual IAT appears to distinguish it from the two other prejudice IATs. This difference suggests that IAT measures with few exemplars in each category are less susceptible to this confound. Should this prove to be a general rule, and if IAT constructs can be assessed reliably with just two (or few) exemplars in each category, a way to create less contaminated IAT measures would be indicated.

While Study 1 showed the existence of a cognitive ability confound on IAT prejudice scores, little is known about its breadth or nature. Because Study 1 examined only how the confound affects IAT-assessed prejudices, its relevance for other IAT constructs is only implied. But if the cognitive confound is endemic to IAT methodology, it should also affect IAT measures of other constructs such as self-esteem (Greenwald & Farnham, 2000). Also, while it is unlikely that the cognitive confound would be specific to the moral – immoral categories used with the prejudices in Study 1, the positive – negative categories are used more commonly, so it is important to examine whether IATs using these categories are affected similarly.

## STUDY 2

### AIMS OF STUDY 2

Given these issues, the aims for Study 2 were twofold. The first was one of replication and generalization, to examine if the same effects are repeated and if the same cognitive confound affects a different IAT measure, namely self-esteem. Also, the categories of moral – immoral used to assess the prejudices in Study 1 were replaced with the positive – negative categories more commonly used. Second, as the results with the antihomosexual IAT suggest, IATs with fewer exemplars in each category may be less susceptible to the cognitive confound than those with more exemplars. By

using either two or eight exemplars for anti-Black IAT (e.g. two or eight stereotypical Black and White names) and self-esteem IAT, as well as by replicating the two-term antihomosexual IAT, we hoped to determine if this is indeed a general rule.

## METHOD

*Materials.* The five IATs used in Study 2 included anti-Black and antihomosexual prejudices, self-esteem, and the two control IATs. The anti-Black and self-esteem IATs were varied so that half of the participants took the anti-Black version with just two exemplars and the self-esteem IAT with eight exemplars; the other half took the eight-exemplar anti-Black and the two-exemplar self-esteem IATs. The antihomosexual IAT again had just two exemplars each for homosexual and heterosexual. For these three measures, eight exemplars drawn from Greenwald and Farnham's (2000) positive-negative categories were used rather than the moral-immoral exemplars used in Study 1. Self-esteem was measured in an ideographic fashion, with each student allowed to select appropriate exemplars for the me-not me categories. For the two-term version, first name and occupation (usually "student") were used. For the eight-term version, students entered their first and last names, occupation, nationality, home state, university, gender, and religion. Exemplars in these same categories that did not represent the student were placed in the not-me list. The control IATs were unchanged from Study 1. Finally, to examine the relationships between IAT and self-report measures, Rosenberg's (1965) Self-Esteem Scale and the anti-Black and antihomosexual scales used in Study 1 were administered.<sup>2</sup>

---

2. The cognitive requirements of the IAT and classic Stroop Color-Word Interference Test (Stroop, 1935), on which individuals must say as quickly as possible the color of a color word written in a different color (e.g., BLUE written in red ink) differ in several ways, but both require participants to suppress the effects of incongruent information in order to correctly respond to the target information. For that reason, we explored in Study 2 if IAT scores correlate with scores on the Stroop test. However, scores on the IATs and Stroop tests were, at best, weakly related. The correlations between the two were significant for just two of seven possible tests. Performance on the Stroop test correlated with the antihomosexuality IAT,  $r = .29, p < .01$ , and with the delicious-happy IAT,  $r = .24, p < .03$ , but its correlations with the flowers-insects IAT and with both versions of the anti-Black and self-esteem IATs did not approach significance. The cognitive demands of the two tasks thus appear to be largely independent of one another.

*Participants.* Ninety-six students, virtually all of traditional college age, participated for course credit. The decision rule for dropping participants used in Study 1 was used again and led to dropping 7 participants. The final sample consisted of 64 females and 25 males; of these, the six Black participants were dropped from the anti-Black analyses but kept for the remaining analyses.

*Procedure.* Participants were again run individually. For each IAT, the number of practice and test trials and their sequence of presentation were the same as in Study 1. Although order effects were not expected, these were controlled as follows: The five IATs were divided into three subsets (Anti-Black with flowers-insects, delicious-not delicious with antihomosexual, and self-esteem). These three subsets yielded six possible orders of IAT presentations, and each order was presented twelve times. Also, the IATs (always given together), the Stroop test (see footnote 2), and the explicit questionnaire also yielded six orders that were rotated in the same fashion. With the six IAT orders embedded within six orders of the other materials, the total set of materials were presented in 36 different orders.

## RESULTS

Order effects were minimal. For the IATs and explicit questionnaires, only one effect was found. Scores on the antihomosexuality IAT were greater,  $F(1, 86) = 6.54, p < .05$ , when the IATs followed the explicit questionnaires. Within the rotation of the IATs, self-esteem scores on the two-exemplar self-esteem IAT were significantly higher in the three orders in which it preceded rather than followed the anti-Black IAT,  $F(1, 41) = 17.81, p < .01$ . Because this order effect did not affect the magnitude of the correlation between the two-exemplar self-esteem IAT and the control IATs, it is not germane to the purposes of this study. No other order effects were found.

Presented in Table 3, the latencies on all IATs were significantly longer with the incongruent presentations, yielding large Cohen's-*ds*. The three IATs replicated from Study 1 yielded virtually identical statistics, except that the delicious IAT had a higher alpha in the second study (Table 2). Of particular importance, however, are the similarities of the IATs for the eight- and

two-exemplar versions of the anti-Black and self-esteem IATs. As shown in Table 3, the two anti-Black IATs had comparable latencies and *ds*. The eight-exemplar version of the self-esteem IAT had longer mean latencies and a slightly larger Cohen's-*d* than the two-exemplar version, but the latter still yielded a substantial *d*. And as shown in the lower half of Table 2, the alpha coefficients for the two-exemplar anti-Black and self-esteem IATs were above .80. Also, the two-exemplar IATs did not lead to systematic range restriction: The long version of the anti-Black IAT had a standard deviation of 181.5 compared to the short versions' 223.1; for self-esteem, the standard deviations were 265.1 and 222.7 for the long and short versions, respectively.

As found in Study 1 and by McFarland and Mattern (2002), the two IAT prejudices were correlated. The eight-exemplar anti-Black IAT correlated .45,  $p < .01$ , with the antihomosexual IAT, although the two-exemplar anti-Black and antihomosexual IATs were not significantly correlated,  $r = .19, p > .10$ . Controlling for both control IATs reduced the first correlation to  $r = .34, p < .02$ ; the latter was reduced to .18. The self-esteem IAT did not correlate significantly with the eight- or two-exemplar version of either prejudice IAT.

As in Study 1, to assess how strongly cognitive skill determines the difference between the congruent and incongruent presentations, repeated measures ANCOVAs were run on each prejudice and self-esteem IAT controlling the IATs used to assess the general cognitive confound. Table 3 shows that the adjusted mean difference between the congruent and incongruent presentations were always reduced, as was the Cohen's-*d* in each case. However, comparing the means and Cohen's-*ds* for eight- and two-exemplar versions of the anti-Black and self-esteem IATs indicates that the reduction is substantially greater for the eight-exemplars measures. However, the substantial reductions for the two-exemplar antihomosexuality IAT (see Table 3) suggests caution in interpreting this difference.

*Confounds on Long vs. Short IATs.* For Study 2, the delicious-happy and flowers-insects IATs correlated .37,  $p < .01$ , so their standardized scores were again summed as an overall index of the cognitive confound. The last half of Table 2 presents the correlations of the prejudice and self-esteem IATs with the two IATs

TABLE 3. Mean Millisecond Latencies for the Five IAT Measures With “Matched” and “Mismatched” Evaluative Exemplars, Study 2

Category matches	Latency (milliseconds)	Cohen's <i>d</i>
Antihomosexuality		
Heterosexual–Positive and Homosexual–Negative	794 <sup>a</sup> [970] <sup>a</sup>	1.64
Heterosexual–Negative and Homosexual–Positive	1146 [1054]	[.53]
Anti–Black Racism (Eight Exemplar Version)		
White–Positive and Black–Negative	742 <sup>a</sup> [914] <sup>a</sup>	1.95
White–Negative and Black–Positive	1086 [1036]	[.73]
Anti–Black Racism (Two Exemplar Version)		
White–Positive and Black–Negative	725 <sup>a</sup> [896] <sup>a</sup>	1.50
White–Negative and Black–Positive	1068 [1050]	[.92]
Self–Esteem (Eight Exemplar Version)		
Me–Positive and Not_Me – Negative	772 <sup>a</sup> [988] <sup>a</sup>	1.90
Me–Negative and Not_Me Positive	1204 [1068]	[.48]
Self–Esteem (Two Exemplar Version)		
Me–Positive and Not_Me – Negative	689 <sup>a</sup> [884] <sup>a</sup>	1.68
Me–Negative and Not_Me Positive	999 [1039]	[1.38]
Taste		
Delicious–Happy and Not Delicious–Unhappy	726 <sup>a</sup>	3.42
Delicious–Unhappy and Not Delicious–Happy	1363	
Flowers–Insects		
Flowers–Pleasant and Insects–Unpleasant	734 <sup>a</sup>	1.69
Flowers–Unpleasant and Insects–Pleasant	998	

<sup>a</sup>The two means differ,  $p < .001$ , on one–sample  $t$ –test.

used to test for the cognitive confound. Both the eight–exemplar anti–Black and eight–exemplar self–esteem IATs correlated substantially with both control IATs and with their standardized sum. Corrected for attenuation, these correlations were quite large, accounting for 53% and 30% of the variance in the anti–Black and self–esteem measures, respectively. However, neither the two–exemplar anti–Black nor two–exemplar self–esteem IAT correlated with either the “delicious” or “flowers” IAT or with their sum. However, the antihomosexual IAT, a two–exemplar IAT, correlated significantly with the delicious IAT and with the sum.



As in the first study, the confound was associated primarily with responses in the incongruent categories. For the eight-exemplar anti-Black IAT, the sum of two control IATs correlated strongly with performance in the incongruent presentations,  $r = .57, p < .001$ , but was less related to performance when presentations were congruent,  $r = .27, p = .05$ . Similarly, for the eight-exemplar self-esteem IAT, the confound correlated with the incongruent presentations,  $r = .42, p < .02$ , but not with the congruent presentations,  $r = -.12, ns$ .

If a common latent construct underlies the IAT and self-report measures of each construct, (a) two-exemplar IATs, because they are less confounded with the cognitive skill, should correlate more with self-report scales than do eight-exemplar IATs, and (b) controlling for the confound should enhance the correlation between the IAT and self-report measures of the same construct. In keeping with this logic, the two-exemplar IAT self-esteem scale correlated significantly with the Rosenberg self-esteem scale,  $r = .28, p = .03$ , but the eight-exemplar version did not,  $r = -.03, ns$ . However, the latter correlation was not enlarged by controlling the cognitive confounds. Also, neither the two- nor eight-exemplar anti-Black IATs correlated significantly with the anti-Black scale (unlike Study 1), nor were their partial correlations significant.

## DISCUSSION

The results of Study 2 extended the cognitive skill confound effects of Study 1 in four ways. First, the use of positive-negative categories showed that the confound was not limited to the moral – immoral categories used in the first study. Second, Study 2 showed that a self-esteem IAT is also affected by the cognitive confound. Third, the Study 1 result that eight-exemplar IATs were affected more by the cognitive confound than the two-exemplar IATs was found to apply to anti-Black and self-esteem IATs as well. The one exception to this general rule was that the antihomosexual IAT in Study 2 correlated significantly with one of the two control IATs. Finally, the finding that the confound affected responses primarily when psychologically incongruent categories were presented was replicated for the anti-Black IAT and extended to the self-esteem IAT, although the confound also

marginally affected responses in the congruent anti-Black IAT. However, controlling the cognitive confounds had little effect upon the correlations between the IAT and self-report assessments of the same constructs.

Importantly, the two-exemplar IATs all had good internal consistency, variance, and *d* scores, and the correlations between the two- and eight-exemplar prejudice measures indicate that they tap the relevant constructs. Taken altogether, the results of Study 2 indicate that two-exemplar IATs are functional and are generally freer of the cognitive ability confound.

### STUDY 3

Across studies 1 and 2, as presented in Table 2, eight-exemplar IATs correlated significantly with the two control IATs on all eight possible occasions; in contrast, the two-exemplar versions did so on just one of eight tests. But before recommending two-exemplar IATs, it is important to know if longer and shorter (e.g., eight vs. two exemplars) versions of the same IAT are correlated, and if they remain so once the variance associated with the control IATs is removed. If not, the two versions do not assess the same construct. Because neither earlier study correlated short and long versions of the anti-Black or self-esteem IATs, Study 3 examined these correlations.

### METHOD

The correlation between two- and eight-exemplar versions of the anti-Black IAT was examined in a sample of 30 students; that for the two- and eight-exemplar self-esteem IATs was inspected on a second sample of 34 students.<sup>3</sup> The two control IATs used in the earlier studies were also administered to each sample. For each sample, the long and short versions were each administered first half of the time, and the two versions were always separated by one of the control IATs.

---

3. Both the anti-Black and self-esteem IATs were given to the first sample, but a programming error on the self-esteem IATs necessitated the second sample.

## RESULTS

Our decision rule resulted in dropping four participants from the first sample. With these participants deleted, the eight- and two-exemplar anti-Black IATs correlated  $.70, p < .001$ . The delicious – happy control IAT correlated  $.46, p < .02$ , with the eight-exemplar anti-Black scale, but the remaining correlations between the anti-Black and control IATs were not significant. Partial correlations removing the control IAT variance changed the correlation between the eight- and two-exemplar anti-Black IATs to  $.71, p < .001$ .

For the second sample, the decision rule again removed four participants. The eight- and two-exemplar self-esteem IATs correlated  $.39, p < .03$ . The delicious – happy IAT also correlated with the eight-exemplar self-esteem IAT,  $r = .36, p < .04$ , but no other control – self-esteem correlation was significant. The partial correlation between the two self-esteem IATs remained significant with the variance of the control IATs removed,  $r = .37, p < .05$ . The anomaly of Study 3 was that the delicious–happy IAT correlated with the two-item anti-Black IAT,  $r = .42, p < .05$ .

## DISCUSSION

In summary, the eight- and two-exemplar versions of the both anti-Black IAT were correlated strongly, and the two self-esteem IATs correlated significantly. Both correlations remained significant with the control IAT variance removed by partial correlation. The strong correlation between the two anti-Black IATs corrected for attenuation was  $.88$ , indicating that little true common variance is lost in using the two-exemplar version. The smaller correlation between the two self-esteem IATs,  $.49$  corrected for attenuation, makes the benefits of using the shorter version less certain. While the two-exemplar version was again uncorrelated with either control IAT, the two self-esteem IATs may share less common variance and may not be as interchangeable as are the anti-Black IATs.

## STUDY 4

Studies 1 and 2 found that the cognitive confound was specifically associated with information processing deficits when terms were

presented in incongruent pairings. However, in both these studies, the congruent pairings for all IATs were presented first, followed by the incongruent pairings. This perfect confounding of order with congruency leaves open the possibility that the switching from the first to the second pairings created the cognitive difficulty rather than the incongruence itself. Study 4 was conducted to examine this possibility.

## METHOD

Sixty-five undergraduate psychology students responded to the eight-exemplar versions of the anti-Black and self-esteem IATs and to the two control IATs. For the anti-Black and self-esteem IATs, 34 participants received the congruent before the incongruent pairings while the other 31 received the incongruent pairings first. On the control IATs, the congruent pairings were always presented before the incongruent pairings. A more complete design would have varied the congruent–incongruent order for the control IATs independently, but varying this order on the test IATs was sufficient to test whether the confound was due to incongruence or order switching. The four IATs were presented in four orders, with each of the two control and two test IATs being presented in each position approximately  $\frac{1}{4}$  of the time. The anti-Black and self-esteem IATs were always separated by one control IAT.

## RESULTS

Our decision rule led to dropping one participant from each subgroup. The two control IATs were correlated  $.46, p < .01$ , so their standard scores were added for an overall cognitive confound control. As shown in Table 4, the incongruent pairings presented first—thereby unconfounded with order switching—correlated with the control IATs on all six possible tests (i.e., correlations in the columns labeled "Inc1"). By contrast, the congruent presentations presented first did not correlate with the control IATs on any of the six tests (correlations in columns labeled "Con1"). Clearly, then, the cognitive skill assessed by the control IATs is

largely associated with processing exemplars presented in incongruent categories. However, as shown in the Table 4 "Con2" columns, the control IATs and congruent pairings were generally correlated when the congruent pairings were presented after the incongruent pairings; these correlations indicate that the cognitive skill is also associated with the difficulty of processing exemplars when category pairings are switched. Finally, as in studies 1 to 3, the control IATs and incongruent pairings were generally correlated when the incongruent pairings were presented after the congruent pairings (the Inc2 correlations). The anomaly from this pattern was that, for unknown reasons, the Flowers–Insects control IAT did not correlate with the incongruent pairings presented in the second position for either the anti-Black or self-esteem IAT.

## DISCUSSION

Despite the anomaly, the conclusion seems warranted that *both* incongruence and the difficulty in mentally switching orders in the middle of an IAT are associated with the cognitive confound. The general cognitive skill assessed by the control IATs predicted speed of responding to incongruent pairings, regardless of their order of presentation. However, this skill was related to congruent pairings when they were presented after the incongruent pairings and required the participants to undo the implicit associations acquired during the first trials. Some respondents had greater difficulty in doing so than others, and this individual difference was also related to the general cognitive skill assessed by the control IATs.

## GENERAL DISCUSSION

Implicit prejudice and other implicit attitudes almost certainly exist, and the IAT is likely to remain a useful tool for their study. However, every measurement method yet devised has systematic method variance, and it would be surprising if the IAT were an exception. As traditionally administered, with many exemplars in each category, IAT scores are confound by a respondent's skill in responding on control IATs. The confound applies to both prejudice and self-esteem IATs and occurs whether the categories of

TABLE 4. Correlations of the Anti-Black and Self-Esteem Congruent and Incongruent Pairings with the Control IATs, Controlling for Order of Presentation

	Anti-Black IAT				Self-Esteem IAT			
	Con1	Inc1	Con2	Inc2	Con1	Inc1	Con2	Inc2
Control IAT								
Delicious (D)	-.01	.76**	.51**	.39*	.07	.33*	.29***	.29***
Flowers (F)	-.11	.48**	.41**	-.01	-.07	.48**	.24	.14
D + F	-.05	.74**	.50**	.24***	.13	.46**	.31*	.25***

Note. Con1 = Congruent pairings presented first; Inc1 = Incongruent pairings presented first; Con2 = Congruent pairings presented following incongruent pairings; Inc2 = Incongruent pairings presented following congruent pairings. \*\* $p < .01$ ; \* $p < .05$ ; \*\*\* $p < .10$ .

moral vs. immoral or positive vs. negative are paired with the tested construct. The magnitude of the confound is substantial, with shared variance ranging in these studies from 17% (for self-esteem IAT, Study 2) to 53% (anti-Black IAT, Study 2), although study 4 indicates that a portion of this shared variance in studies 1 to 3 is attributable to order switching rather than category incongruence.

The implications of this confound are straightforward: Participants with little difference in their response speeds to incongruent versus congruent pairings on control IATs are biased toward lower IAT prejudice scores; those with larger differences are biased toward higher IAT-assessed prejudice. For self-esteem IATs, those affected little by category incongruence versus congruence are biased toward higher self-esteem scores, while those affected substantially are biased toward lower self-esteem.

While these studies examined only prejudice and self-esteem IATs, the results are sufficient to suggest that this cognitive confound is endemic to IAT methodology. It is likely to affect virtually all IATs as they are now used. Just as researchers who use self-report measures must attend to social desirability and response set confounds, those who use IATs in future studies need to be mindful of its cognitive ability confound. The confound is especially problematic for researchers who choose to examine relations between or among IATs; the common method artifact is likely to inflate a weak relationship or indicate a significant relationship where none actually exists. The surge of new IATs and of studies using IAT methodology in the last few years magnifies the importance of recognizing and controlling this confound.

Statistically controlling this confound reduced, but did not eliminate, the correlations among IAT-assessed prejudices. On the other hand, neither controlling the confound nor using two-exemplar IATs appreciably enlarged the correlations between the IAT and self-report measures of prejudice or self-esteem. The failure of controlling for the cognitive confound to enlarge the cross-method correlations of the same constructs may simply mean that the two measures do not reflect a common latent construct. Despite this failure, in situations where IAT measures predict specific behaviors, controlling for the cognitive confound seems likely to enhance its correlations with the behaviors.

Two methods could be used for controlling it. First, investigators could administer control IATs and partial out variance associated with them. Certainly the two used here together extract a mass of variance associated with this confound, and the added time required to administer both is about ten minutes. With IAT tests already developed, and where researchers want to compare new findings with older results, this method might be appropriate.

However, the current results indicate that IATs can easily be written that are largely free of this confound. The two-exemplar IATs examined here (antihomosexual, anti-Black, and self-esteem) retained good internal consistency and variance, and were significantly correlated with the two control IATs for only two of ten tests across studies 1 through 3. Further, the significant correlations of the long and short prejudice IATs with each other in Study 3 suggests that variance associated with the assessed constructs is retained.

One objection to using two-exemplar IATs might be that participants' specific associations with the chosen exemplars are more likely to influence their IAT scores. That is, in responding to the categories Black versus White, personal positive associations with the names Nancy (perhaps one's favorite aunt), or negative ones with Latisha, could mean that a participant's IAT score reflects associations with particular names rather than with the categories. Embedded within more exemplars, personal associations with particular names would have less influence. However, De Houwer (2001), using a British – foreign, positive – negative IAT, showed that the target concept rather than the chosen exemplars appears to determine IAT results. Using three exemplars each of

positive and negative British names (e.g. Princess Diana, mass murderer Donald Shipman, respectively) and positive and negative foreign names (Ghandi, Hitler), De Houwer found that the categories of British versus foreign drove IAT responses and that the positive versus negative exemplars had almost no effect upon performance. Given the strength of the positive and negative associations with the names tested by De Houwer, individuals' associations with exemplars in two-exemplar IATs probably affect IAT results very little.

We can only speculate as to why IATs with just two exemplars in each category of the tested construct are so little affected by this cognitive confound. It seems likely, however, that the cognitive requirements of retaining awareness of the incongruent categories yet suppressing their incongruence constitutes, in itself, a heavy cognitive load. In this context, the use of fewer exemplars may enable participants to focus more specifically on the construct being tested. In contrast, longer lists of exemplars may increase the cognitive load and thereby reduce the respondent's available capacity for actively blocking category incompatibility. Creative investigations of the cause of this difference between two- and many-exemplar IATs are certainly warranted.

Whatever its precise cause, the cognitive skill that affects IAT responses may have systematic effects upon particular populations. Elderly individuals, for example, have greater difficulty on several measures that require suppressing incongruent information (Kane, Hasher, Stoltzfus, Zacks, & Connelly, 1994), including the Stroop test (Uttl & Graf, 1997) and visual search tasks (Madden, 1983), and story-reading-with-distractions task (Connelly, Hasher, & Zacks, 1991). Despite the unique cognitive requirements of the IAT, its common requirement with other tasks that demand suppressing incongruent information suggests that older individuals may be biased toward greater prejudice and lower self-esteem on IAT measures of these constructs. Similarly, those with identifiable attention disorders such as schizophrenia (Frith, 1979) and ADHD (MacLeod & Prior, 1996) have greater difficulty on tasks that require suppression and may be adversely affected by the cognitive requirements of the IAT. To date, the effect of age, attention disorders, and other individual differences upon IAT scores have not been examined.



**APPENDIX  
EXEMPLARS USED IN THE IMPLICIT ASSOCIATION TESTS**

**Control IATs**

**1. Delicious**

Candy  
Yummy  
Strawberry  
Tasty  
Cookies  
Chocolate

**Not Delicious**

Liver  
Spoiled  
Rancid  
Turnips  
Rotten  
Anchovies

**Happy**

Joy  
Laughter  
Gleeful  
Optimism  
Merry  
Cheerful  
Blissful

**Unhappy**

Depressed  
Despair  
Gloom  
Pessimism  
Sobbing  
Misery  
Hopeless

**2. Flower**

Rose  
Daisy  
Tulip  
Orchid  
Iris  
Bluebell  
Pansy  
Lily

**Insects**

Ant  
Flea  
Spider  
Fly  
Termite  
Mosquito  
Moth  
Roach

**Pleasant**

Friend  
Cheer  
Gift  
Love  
Vacation  
Lucky  
Hug  
Sunshine

**Unpleasant**

Filth  
Sickness  
Accident  
Pollution  
Jail  
Cancer  
Vomit  
Stink

**Exemplars used in Prejudice IATs**

**Black**

Latonya  
Shavonn  
Tashika  
Ebony  
Tameka  
Shereen  
Latisha  
Sharonda

**White**

Heather  
Nancy  
Mary  
Margaret  
Melanie  
Stephanie  
Catherine  
Rebecca

**American**

New York  
Baseball  
Thanksgiving  
Hollywood  
Cowboy  
Hamburger  
Yankee  
Washington  
Buffalo  
Jeans

**Foreign**

China  
Sombrero  
Pyramids  
Burritos  
Hindu  
Gandhi  
Jungle  
Gorbachev  
Jungle  
Chopsticks

<b>Homosexual</b>	<b>Heterosexual</b>
Gay	Straight
Lesbian	Married
<b>Moral</b>	<b>Immoral (used with all prejudice categories in Study 1)</b>
Honest	Deceiving
Helpful	Selfish
Kind	Cruel
Loving	Hateful
Thoughtful	Abusive
Generous	Stingy
Sincere	Negligent
Responsible	Lying
<b>Positive</b>	<b>Negative (used with prejudice categories in Study 2 – 4)</b>
Joy	Stupid
Smart	Abuse
Nice	Sick
Smile	Mean
Loved	Hatred
Peace	War
Honest	Liar
Health	Agony

## REFERENCES

- Altemeyer, B. (1988). *Enemies of freedom*. San Francisco: Jossey-Bass.
- Asendorph, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83*, 380–393.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Connelly, S. L., Hasher, L., & Zacks, R. T. (1991). Age and reading: The impact of distraction. *Psychology and Aging, 6*, 533–541.
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12*, 163–170.
- De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology, 37*, 443–451.
- De Jong, P. J., Pashman, W., Kindt, M., & van den Hout, M. A. (2001). A reaction time paradigm to assess (implicit) complaint-specific dysfunctional beliefs. *Behavior Research & Therapy, 39*, 101–113.
- Egloff, B., & Schulke, S. C. (2002). Predictive validity of an implicit association test for assessing anxiety. *Journal of Personality and Social Psychology, 83*, 1441–1455.

- Forsyth, R. A. & Feldt, L. S. (1969). An investigation of empirical sampling distributions of correlation coefficients corrected for attenuation. *Educational and Psychological Measurement, 29*, 61–71.
- Frith, C. D. (1979). Consciousness, information processing, and schizophrenia. *British Journal of Psychiatry, 134*, 225–235.
- Gemar, M. C., Segal, Z. V., Sagrati, S., Kennedy, S. J. (2001). Mood-induced changes on the Implicit Association Test in recovered depressed patients. *Journal of Abnormal Psychology, 110*, 282–289.
- 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*, 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*, 1464–1480.
- Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie, 48*, 85–93.
- Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language, 45*, 82–102.
- Kane, M. J., Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Connelly, S. L. (1994). Inhibitory attentional mechanisms and aging. *Psychology and Aging, 9*, 103–112.
- Kite, M. E. & Deaux, K. (1986). Attitudes toward homosexuality: Assessment and behavioral consequences. *Basic and Applied Social Psychology, 7*, 137–162.
- MacLeod, D., & Prior, M. (1996). Attention deficits in adolescents with ADHD and other clinical groups. *Child Neuropsychology, 2*, 1–10.
- Madden, D. J. (1983). Aging and distraction by highly familiar stimuli during visual search. *Developmental Psychology, 19*, 499–507.
- McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner. *Prejudice, discrimination, and racism*. New York: Academic Press.
- 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*, 435–442.
- McFarland, S. G. & Mattern, K. (2002). *Generalized Explicit and Implicit Prejudice*. Manuscript submitted for publication.
- Nunnally, J. C. (1967). *Psychometric Theory*. McGraw Hill.
- 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*, 97–144.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, N.J.: Princeton University Press.
- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57*, 743–762.

- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–661.
- Teachman, B. A., Gregg, A. P., Woody, S. R. (2001). Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology, 110*, 226–235.
- Uttl, B., & Graf, P. (1997). Color-Word Stroop test performance across the adult life span. *Journal of Clinical and Experimental Neuropsychology, 19*, 405–420.