The Brief Implicit Association Test

N. Sriram[1] and Anthony G. Greenwald[2]

[1]University of Virginia, Charlottesville
[2]University of Washington, Seattle

Abstract. The Brief Implicit Association Test (BIAT) consists of two blocks of trials with the same four categories and stimulus-response mappings as the standard IAT, but with 1/3 the number of trials. Unlike the standard IAT, the BIAT focuses the subject on just two of each block's four categories. Experiments 1 and 2 demonstrated that attitude BIATs had satisfactory validity when *good* (but not *bad*) was a focal category, and that identity IATs had satisfactory validity when *self* (but not *other*) was a focal category. Experiment 2 also showed that a *good*-focal attitude BIAT and a *self*-focal identity BIAT were psychometrically similar to standard IAT measures of the same constructs. Experiment 3 presented each of six BIATs twice, showing that procedural variables had no more than minor influences on the resulting implicit measures. Experiment 4 further demonstrated successful use of the BIAT to measure implicit stereotypes.

Keywords: Implicit Association Test, implicit measure, social cognition

In eleven years since its introduction, the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) has been used in several hundred studies to provide measures of association strengths. The associations investigated have often corresponded to *attitudes*, *identities*, and *stereotypes* (Greenwald et al., 2002). Attitude IATs combine a concept classification (e.g., *Coke* vs. *Pepsi*) with an attribute classification representing positive versus negative valence (e.g., *pleasant* vs. *unpleasant*). Identity (or self-concept) IATs combine contrast of *self* versus *other* with a nominal contrast (e.g., *male* vs. *female*; *family* vs. *career*; *math* vs. *arts*) or a trait contrast (e.g., *strong* vs. *warm*; *large* vs. *small*). Stereotype IATs combine social group categories (e.g., *male* vs. *female*; *Asian* vs. *Hispanic*) with nominal or trait contrasts.

In *combined task* blocks of the IAT, subjects switch between classifying exemplars of one contrast (e.g., *Pepsi* vs. *Coke*) and exemplars of the other contrast (e.g., *pleasant* vs. *unpleasant*) on consecutive trials. In a {*pleasant*+*Pepsi*}|{*unpleasant*+*Coke*} combined task, *pleasant* and *Pepsi* are mapped to one response (e.g., left key) and *unpleasant* and *Coke* to the other response (e.g., right key). A second combined-task block reverses the response mappings of one of the contrasts (e.g., {*pleasant*+*Coke*}|{*unpleasant*+*Pepsi*}). If *Pepsi* is more strongly associated with positive valence than *Coke*, classification should be faster in the {*pleasant*+*Pepsi*}|{*unpleasant*+*Coke*} block than in the {*pleasant*+*Coke*}|{*unpleasant*+*Pepsi*} block.

Various strategies can be used in performing the IAT's combined tasks. One strategy is to prepare equally for all four of the category–response mappings (e.g., *pleasant*–left; *Pepsi*–left; *unpleasant*–right; *Coke*–right). Alternately, subjects can focus on just the two mappings associated with (say) the left-side response, giving themselves an added mental note to give the right-side response for "anything else". Although such strategy variations have not been the focus of published research, designed variations on IAT procedures (e.g., the Go/No-go Association Task of Nosek & Banaji, 2001 and the single-category IATs of Bluemke & Friese, 2008; Karpinski & Steinman, 2006) have had at least the partial intent of managing the subject's performance strategy. The present research investigates a new modification of the IAT – a Brief Implicit Association Test (BIAT) – that uses simplified instructions and was designed with the intention of reducing spontaneous variation in subject strategy.

The IAT has four categories, each mapped onto one of the two responses. The BIAT instructs respondents to focus on just two of the IAT's four category–response mappings in each combined task. Prior to each combined task, subjects are shown two category labels together with their exemplars and are instructed (a) to "keep them in mind", (b) to respond to items from these two categories with a "focal" response key, and (c) to respond to any other stimuli with an alternative "nonfocal" response key. With two blocks and a total of fewer than 80 trials, the BIAT substantially reduces administration time relative to the standard five-step IAT procedure, which is often done with seven blocks of trials and typically involves 180 or more trials.

The four experiments in this report investigate properties of the BIAT. Experiment 1 provides initial assessments of the BIAT's psychometric properties, unexpectedly finding that it matters which subset of the four category–response mappings is selected for focus. Experiment 2 establishes convergence between BIAT and standard IAT measures of attitude and identity. Experiment 3 extends the BIAT to additional attitude and identity topics, and also to stereotypes. Experiment 4 focuses on implicit stereotypes, including race, age, and gender stereotypes. Collectively, the four experiments establish the ability of the BIAT to function effectively in the range of domains in which standard IAT measures have been successfully used.

General Method

Subjects

Subjects were undergraduate students from the University of Washington Psychology Department's undergraduate subject pool, who provided data at desktop computers in individual cubicles.

Design of the BIAT

Each BIAT is composed of two combined-task blocks, each of which can be described by its two focal categories (e.g., a block with *pleasant* and *Pepsi* focal might be followed by a block with *pleasant* and *Coke* focal). One category is focal in both combined tasks (*pleasant* in the example just given) and has a contrasting category that remains nonfocal in both tasks (*unpleasant* in this example). Our convention for naming BIATs lists the four categories, placing the category that consistently remains nonfocal within parentheses. The implicit soft-drink BIAT measure in this example is named *Coke–Pepsi/pleasant–(unpleasant)*. As a further convention, the order of listing indicates interpretation of scores. High scores indicate greater strength of the association of the first-listed category with the third. In the *Coke–Pepsi/pleasant–(unpleasant)* BIAT, scores above zero indicate that the *Coke–pleasant* association is stronger than the *Pepsi–pleasant* association. Table 1 describes a general schema for the structure of BIAT measures, illustrated for this soft-drink attitude example.

Procedure

Prior to completing BIATs, subjects completed parallel self-report measures of strength of the corresponding associations. The instructions for each BIAT block displayed all exemplars for the upcoming block's two focal categories (see Figure 1). The two focal categories are typically distinguished from each other not only by category identity but also by visual format (e.g., text vs. image, or distinct fonts if both are textual). Subjects required an average of about 10 s to process the BIAT block instructions.

After the instruction page display (e.g., Figure 1), the lists of focal-category exemplars disappeared, but the focal category labels remained in view. On each BIAT trial, an exemplar of one of the four categories appeared in center screen. If the initial response to a stimulus was in error, subjects were obliged to give a second response, and latency was recorded to occurrence of the correct response. This created a built-in error penalty, which is also a property of standard IAT measures (cf. Greenwald, Nosek, & Banaji, 2003).

The interval between the correct response on one trial and presentation of the next stimulus was 400 ms.[1]

Response errors were signaled by a red "X", which appeared centered below the stimulus and disappeared immediately when the correct response was made. The studies were administered in individual subject stations using Inquisit Version 2.0 (2005) to control computer displays and data recording. At the conclusion, subjects received on-screen debriefing information.

Analysis Strategy

Measures of association strength based on IATs and BIATs were computed using the *D* measure (Greenwald et al., 2003), which is an effect-size-like measure with possible range of -2 to $+2$. *D* is computed as the difference between mean latencies of the two BIAT blocks divided by the *inclusive* (not pooled) standard deviation of latencies in the two blocks. This measure has been shown to have psychometric properties superior to those of a wide variety of alternative strategies for using latencies from the IAT's two tasks (Greenwald et al., 2003; Sriram, Nosek, & Greenwald, 2007).[2]

To estimate internal consistency of BIAT measures, split-half reliabilities were computed by partitioning the trials in each of the two blocks into two parallel subsets. For example, in the 32-trial blocks used in Experiments 1 and 2, one subset consisted of trials {1, 2, 5, 6, 9, 10, 13, 14, 19, 20, 23, 24, 27, 28, 31, 32} and the other subset consisted of the remaining 16 trials. In the 20-trial blocks of Experiments 3 and 4, one subset had trials {5, 6, 9, 10, 15, 16, 19, 20} and the other subset was composed of trials {7, 8, 11, 12, 13, 14, 17, 18}; the practice, prefatory trials 1–4 were unanalyzed. *D* measures for each subset were computed from the differences between mean latencies of the same subset in the two blocks, divided by the inclusive standard deviation of these latencies. Internal consistency was estimated as the correlation between these split halves. Each BIAT was administered twice, permitting computation of a *test–retest* estimate of reliability.

Explicit attitude measures included (a) the difference between separately rated strengths of association of contrasted concepts with positive or negative valence (e.g., difference between the liking ratings for cola brands) and (b) single-item Likert-format measures of relative preference between the contrasted concepts. Similar combinations based on sets of three items were used to obtain measures of relative strength of associations of *self* with contrasted identities based on gender and ethnicity. As described by Greenwald et al. (2003), the correlations between implicit and parallel explicit measures served as the primary validity criteria for attitude and identity BIATs. The stereotypes examined in this research were ones that had been demonstrated to be

---

[1]   Variation in this interval (150–750 ms) was previously shown to have no significant impact on IAT measures (Greenwald et al., 1998).
[2]   Values absolutely larger than 1.5 are rarely observed with the *D* measure. Because standard deviations of *D* measures are typically smaller than 0.5, means of *D* measures most often correspond to Cohen's *d* values at least twice as large (see, e.g., the IAT *D*s and their corresponding Cohen's *d*s in Table 5).

Table 1. Structure of BIAT measure for associations with positive valence[a]

Block[b]	No. of trials[c]	Function	General instruction for nonfocal key: "Press for anything else" [items are shown in brackets]	Instruction for focal key[d]
1	20	First combined task	[Focal concept 2[e] or unpleasant words]	Press for focal concept 1[e] or pleasant words
2	20	Second combined task	[Focal concept 1 or unpleasant words]	Press for focal concept 2 or pleasant words
3	20		Repeat Block 1 [alternately, repeat Block 2]	
4	20		Repeat Block 2 [alternately, repeat Block 1]	

[a] Attitude measures with valence as attributes are likely to be the most frequent uses of BIAT measures. On the basis of present findings, it is recommended that assigning the *pleasant* valence attribute to the focal key yields superior measures. As is true of the standard IAT, the BIAT can be done with any concept used as focal in both of the first two blocks and, likewise, any suitable contrasting concept nonfocal in both of these blocks.

[b] Repetitions after the first two blocks are optional. Each pair of blocks produces a separate BIAT measure. Averaging these multiple BIAT measures should increase measure reliability and sensitivity.

[c] Trials 1–4 in each block may be limited to just the two concepts that alternate focal status in Blocks 1 and 2 (as in present Experiments 3 and 4), and then omitted from data analyses. This use of Trials 1–4 reinforces instructions about which of these two concepts is focal in each block. Total numbers of trials in each block can be increased, which should increase reliability and sensitivity of measures.

[d] Arbitrarily, the focal key was always on the right side in the present experiments. Properties of BIAT measures should be unaffected by interchanging right and left key functions.

[e] Focal concepts 1 and 2 are the two concepts that alternate focal status in Blocks 1 and 2. The order in which these concepts are made focal is ordinarily counterbalanced.

{Pepsi}

{PLEASANT}

NICE, HEAVEN, HAPPY, PLEASURE

Two categories, and their items, are displayed above.

Keep the two categories in your mind as you do the task.

Press 'K' when an item matches EITHER category.

Press 'D' for anything else.

If you make an ERROR you will see a RED X.

When this happens, make the CORRECT response to proceed.

Go FAST. A few errors are OK.

Press the Spacebar to begin the task.

Figure 1. Instruction screen for the *{Pepsi+pleasant}* BIAT block.

societally pervasive in previous research (Nosek et al., 2007). Partly because of limited individual-difference variation, implicit–explicit correlations are weaker for these than for the attitudes and identities examined in the present research. For these pervasive stereotypes, the main test of validity of BIAT measures was their ability to detect the same stereotype that had been found in previous research with standard IAT measures.

Data analyses used hierarchical multiple regressions. In the first regression step, the D score, as criterion, was regressed onto counterbalanced experimental design factors and their interactions. In the second step, the parallel self-report measure was added as a predictor. In the third and final step, interactions of the self-report predictor with design factors were added as predictors. For attitude and identity measures, evidence for BIAT validity took the form of significant prediction of the IAT measure by the self-report measure in the second step. Evidence for validity was strengthened if the self-report measure's relationship to the BIAT-measure criterion was not moderated by design factors in the third step.

Experiment 1

Overview

Experiment 1 was conducted shortly before the 2004 US Presidential Election involving George W. Bush and John F. Kerry. An attitude BIAT assessed implicit candidate preference and an identity BIAT measured the association between self and gender. Previous research has shown

substantial implicit–explicit correlations in these domains (e.g., Aidman & Carroll, 2003; Greenwald et al., 2003; Nosek, 2005; Rudman, Greenwald, & McGhee, 2001). These correlations, which were expected to be at least moderate in size, were expected to provide a useful check on the validity of the BIAT format.

Unlike standard IATs, each BIAT has up to four variants that differ on which of the four component categories is never focal in the two combined-task blocks. In the candidate attitude BIATs, two of these four variants were compared – *Kerry–Bush/good–(bad)* and *Bush–Kerry/ bad–(good)*. In the first of these, the category *bad* was never focal, and the measure was scored so that strong associations of *Kerry* with *good* received high scores. The second was scored in the same direction (association of *Kerry* with *bad* received low scores) and the category *good* was never focal. The two gender identity BIATs were identified as *female–male/self–(other)* and *male–female/other–(self)*, both scored so that stronger associations of *female* with *self* than with *other* would receive high scores.

Stimuli

In the *Kerry–Bush/good–(bad)* and *Bush–Kerry/bad–(good)* BIATs, four face images of each presidential candidate were used as category exemplars. Exemplars for *good* were the four words, happy, warm, love, and friend; exemplars for *bad* were angry, cold, hate, and enemy. Stronger associations of *Kerry* with *good* than *bad* received high scores. For the *female–male/self–(other)* and *male–female/other– (self)* BIATs the categories were *female* (female, woman, girl, she), *male* (male, man, boy, he), *self* (I, me, mine, self), and *other* (they, them, their, other). Stronger associations of *female* with *self* than with *other* received high scores.

Design

Each subject completed eight BIATs, consisting of two repetitions in immediate succession of each of the four distinct 64-trial BIATs. Order of combined tasks within BIATs was counterbalanced across subjects. For example, the *Kerry– Bush/good–(bad)* was either ordered as {*good+Bush*} followed by {*good+Kerry*} or as {*good+Kerry*} followed by {*good+Bush*}. Each block had 32 trials. For half the subjects, the *good* and *self* versions preceded the *other* and *bad* versions and the remainder received the reverse order.

Subjects and Procedure

Subjects (24 females and 16 males) were first asked to self-identify as male or female, after which they provided warmth ratings on a 10-point scale (1 = very cold and 10 = very warm) toward each of George W. Bush, John F. Kerry, the Republican Party, and the Democratic Party. They then reported political identity along a 5-point scale anchored by Democrat and Republican. Next, subjects

practiced the BIAT instructions using two nonsocial focal categories – *curved* (circle, oval, ring, ball) and *bird* (eagle, swan, parrot, duck) – in a single 32-trial training block. Nonfocal exemplars were drawn from *angled* (triangle, square, block, pyramid) and *mammal* (elephant, bison, deer, cow). After this training block, each of the four BIATs was administered twice in immediate succession. Subjects received one of 16 counterbalanced task sequences that varied the position of each BIAT within the overall sequence and the order of combined blocks within each BIAT.

Results

Correct responses to focal category items are called *focal* responses; those to nonfocal categories are *nonfocal* responses. Across the four BIATs, focal responses were faster (709 ms) than nonfocal responses (774 ms), $t(39) = 9.6$, $p = 10^{-11}$ and also had fewer errors, 8.5% versus 10.0%, $t(39) = 2.9$, $p = .007$. These findings were consistent with the expectation that subjects would selectively attend to the focal categories.

Candidate Attitude BIATs

Eight subjects had identical warmth ratings for both presidential candidates. Of the remaining 32 subjects, nine accorded Bush greater warmth and 23 did so for Kerry. A measure indicating preference for Kerry over Bush was constructed by subtracting the Bush rating from the Kerry rating (possible range: −9 to +9).

In the *Kerry–Bush/good–(bad)* BIAT, Bush supporters were faster on trials in the {*good+Bush*} block (658 ms) than on trials in the {*good+Kerry*} block (763 ms). Likewise, Kerry supporters were faster on trials in the {*good+ Kerry*} block (741 ms) than in the {*good+Bush*} block (874 ms). The IAT effect (*D* measure) was satisfactory in internal consistency, and the correlation between the self-reported candidate attitude and the average of the two *Kerry–Bush/good–(bad)* BIATs was .76 (see Table 2).

The hierarchical regression analysis of *D* measures for the *Kerry–Bush/good–(bad)* BIAT included in its first step order (of combined-task blocks), half (first or second half of the experiment), and their interaction. This first step revealed an order effect in which implicit preference for Kerry was higher when the block with *good* and *Bush* focal preceded the block with *good* and *Kerry* focal, $t(35) = 2.2$, $p = .04$. The expected strong effect of explicit (self-report) preference emerged clearly in the second step, $t(35) = 6.6$, $p = 10^{-7}$, zero-order $r = .76$. The third step included as predictors the multiplicative products of the explicit measure with order of combined tasks, half of the experiment, and their interaction (cf. West, Aiken, & Krull, 1996). None of these factors qualified the second step's prediction of IAT-effect *D* measures by explicit preference ($ts < 1$).

For the *Bush–Kerry/bad–(good)* BIAT, mean latencies on {*bad+Kerry*} and {*bad+Bush*} blocks were 712 ms and 778 ms for Bush supporters, compared to 842 ms and 810 ms for Kerry supporters. Unexpectedly, neither the

Table 2. Internal consistency and validity in Experiment 1 ($N = 40$)

| BIAT[a] | Consistency and test–retest correlation | | | | Implicit–explicit correlation | | | | |
| | | | | | All | Trial type | | Administration | |
	α	α_1	α_2	r_{12}		Focal	Non	1	2
Kerry–Bush/good–(bad)	.83	.82	.66	.49	.76	.75	.65	.70	.60
Bush–Kerry/bad–(good)	.76	.79	.66	.17	.11	−.02	.22	.05	.12
Female–male/self–(other)	.94	.92	.83	.55	.70	.73	.60	.73	.47
Male–female/other–(self)	.55	.62	.51	.07	.07	−.14	.25	.06	.04

[a] In naming BIATs (see text), the category named third is focal in both combined-task blocks; the category in parentheses is not focal in either combined task.

Note. Internal consistencies α_1 and α_2 are for the two 64-trial BIAT measures that are intercorrelated (r_{12}) and for their combination (α). Validity is the correlation between corresponding implicit and explicit measures. Validity correlations are presented overall (All), across focal and nonfocal trials (focal and non), and for first or second BIAT administrations (1 and 2). For $N = 40$, r values associated with two-tailed α levels of .05 and .01 are .31 and .40, respectively.

test–retest correlation for this BIAT nor the implicit–explicit correlation was statistically significant. Also inconsistent with expectation, the *Kerry–Bush/good–(bad)* and *Bush–Kerry/bad–(good)* BIATs were not significantly intercorrelated, $r = .18$, $p = .28$.

Gender Identity BIATs

In the *female–male/self–(other)* BIAT dichotomous subject gender served as the explicit measure and correlated $r = .70$ with the average D from the two identical administrations of the BIAT measure. Female subjects were faster on {*self+female*} (596 ms) than on {*self+male*} (727 ms), while male subjects were faster on {*self+male*} (722 ms) than {*self+female*} (785 ms). Internal consistencies of the *female–male/self–(other)* BIAT were satisfactory and the test–retest correlation was $r = .55$, $p = .0003$. In the hierarchical regression, the expected main effect of self-reported sex was evident in the second step, $t(35) = 5.1$, $p = 10^{-5}$, zero-order $r = .70$, and was not moderated by design factors in the third step.

In the *male–female/other–(self)* BIAT, mean latencies on {*other+female*} and {*other+male*} blocks were 722 and 750 ms, respectively, for male subjects and were 715 and 690 ms, respectively, for female subjects. Contrary to expectations, for this BIAT, both internal consistency and test–retest correlation for the D measure were low (see Table 2). The D measure was uncorrelated with subject gender and was also unrelated to the *female–male/self–(other)* variant, $r = -.04$. No significant effects emerged from the three-step hierarchical regression.

Discussion

The choice of focal categories had striking and unexpected effects on BIAT measures. Variants that used *good* or *self* as focal categories produced internally consistent and predictively valid implicit attitude and implicit identity measures.

However, those in which *bad* or *other* were focal yielded measures that were psychometrically weak and did not correlate in expected fashion with parallel self-report attitudes and self-reported sex, respectively.

Implications for Theoretical Understanding of IAT Measures

In Rothermund and Wentura's (2004) *salience asymmetry* interpretation of IAT measures, negative valence (e.g., *bad*) is a "figural" category that is cognitively salient in the context of positive valence (e.g., *good*). Similarly, *other* is figural and salient in the context of *self*. Rothermund and Wentura used this theoretical interpretation to suggest that subjects easily give the same response when two salient or two nonsalient categories are assigned to the same response. If the salience asymmetry interpretation is correct, then one would expect that BIATs that instruct focus on the categories assumed to be figural and salient would produce measures superior to BIATs that instruct focus on nonfigural categories. Although that expectation was clearly disconfirmed by Experiment 1's findings, the results nevertheless supported the more general underlying idea that there is an important asymmetry in pairs such as *good–bad* and *self–other*.

Proctor and Cho's (2006) *polarity correspondence* theory resembles the salience asymmetry theory in supposing that correspondence between categories on polarity (a term encompassing salience, familiarity, and linguistic markedness) underlies performance on IAT measures. Remarkably, the definition of salience in Proctor and Cho's polarity correspondence is diametrically opposed to that in the salience asymmetry theory – Proctor and Cho understand *good* and *self* (rather than *bad* and *other*) to be salient. However, they point out that this definition reversal is inconsequential regarding interpretation of the IAT (Proctor & Cho, p. 433, footnote 4) because both theories predict faster responses when salient (or nonsalient) categories share a response than otherwise. These cognitive asymmetries are

likely important in the BIAT; they are considered further in the General Discussion.

Experiment 2

Overview

Experiment 2 assessed convergence between the BIAT measures that were first used in Experiment 1 and corresponding standard seven-block IAT measures. Subjects (22 males and 45 females) provided data for both formats.

Design

The standard IAT is a five-step, seven-block procedure first reported by Farnham and Greenwald (1999; see also Greenwald & Farnham, 2000), as a reduction of the considerably longer procedure used in the first IAT report by Greenwald et al. (1998). In Experiment 2's standard IAT procedure, the first two of these blocks had 16 trials each. They were followed by two identical 32-trial combined-task blocks after which came a fifth 16-trial block that reversed the concept classification. Two 32-trial combined-task blocks that incorporated the reversed concept classification completed the sequence, for a total of 176 trials. The IAT measure obtained from the standard procedure was computed from two latency contrasts, one based on Blocks 3 and 6 and the other on Blocks 4 and 7 (Greenwald et al., 2003). The BIAT measures for Experiment 2 were identical to those of Experiment 1, with two 32-trial combined tasks each, administered twice in succession.

Procedure

Experiment 2 used the same categories and exemplars as Experiment 1. The explicit scales for the political attitude test were converted to a 9-point format (1 = "very cold" and 9 = "very warm") that included a neutral point. The resulting difference measure (Kerry warmth minus Bush warmth) ranged from −8 to +8. Subjects were administered one of eight counterbalanced task sequences. The response keys "K" and "D" were used for the standard IATs, with "K" consistently corresponding to *good* or *self* and "D" to *bad* or *other*. The BIATs used "P" and "Q", respectively, as the keys for focal and nonfocal responses.

After providing self-report measures, half the subjects did one each of the standard candidate attitude and gender IATs followed by the four BIAT variants used in Experiment 1. Each BIAT was administered twice in succession. The remainder completed the BIATs prior to the standard IATs. As in Experiment 1, a 32-trial {curved+bird} practice block preceded the BIATs. The order of combined-task blocks was varied by counterbalancing but was the same for the brief and standard IATs done by the same subject. Within their half, the standard attitude IATs preceded the standard identity IATs (or vice versa). Experiment 1 had shown that the *good*-focal

and *self*-focal variants of the BIAT had greater validity than the *bad*-focal and *other*-focal variants. Experiment 2's primary objective was to compare standard IATs with the more valid BIAT variants. To this end, although both variants were used, the *good*-focal and *self*-focal variants always preceded the *bad*-focal and *other*-focal variants.

Analysis

As in Experiment 1, BIAT latencies were faster on focal than on nonfocal trials (726 ms vs. 777 ms, $t(66) = 7.6$, $p = 10^{-9}$) and were also more accurate (7.8% vs. 8.7% errors, $t(66) = 2.5$, $p = .016$). For the standard IAT, mean latency (703 ms) and mean error rate (7.8%) did not differ between right and left response keys. Twelve subjects had identical warmth ratings for both candidates, 13 subjects had higher ratings for Bush, and 42 expressed greater warmth for Kerry. Relative warmth for Kerry over Bush provided the validity criterion for candidate attitude IATs. Subject gender provided the validity criterion for gender identity IATs.

Candidate Attitude IATs

In the standard attitude IAT and the *Kerry–Bush/good–(bad)* BIAT variant, the patterns of means in the combined blocks for the pro-Bush and pro-Kerry subjects were similar to those reported in Experiment 1. Validity correlations of IAT measures with self-report measures were strongly positive for both the standard attitude IAT and the *Kerry–Bush/good–(bad)* BIAT (see Table 3). Replicating Experiment 1, this validity correlation was considerably lower for the *Bush–Kerry/bad–(good)* variant. The internal consistencies were markedly higher for the standard attitude IAT and the *Kerry–Bush/good–(bad)* BIAT than they were for the *Bush–Kerry/bad–(good)* BIAT.

Hierarchical regressions for the three attitude IAT measures found no effects of design factors or their interactions in the first step. A strong effect of explicit preference emerged in the second step for both the standard IAT, $t(62) = 6.5$, $p = 10^{-7}$, zero-order $r = .65$, and the *Kerry–Bush/good–(bad)* BIAT, $t(62) = 6.8$, $r = .65$, $p = 10^{-8}$, zero-order $r = .63$, Like Experiment 1, the effect of explicit preference was not significant in the hierarchical regression of the *Bush–Kerry/bad–(good)* BIAT, $t(62) = 1.6$, $r = .22$, $p = .13$. The standard IAT correlated $r = .65$ ($p = 10^{-8}$) with the *good*-focal BIAT, but only $r = .39$ ($p = .001$) with the *bad*-focal BIAT. The correlation between the two BIAT variants was $r = .29$, $p = .02$.

Gender Identity IATs

A positive correlation between subject gender and the IAT D measure would show that, as expected, male subjects were faster when *self* and *male* shared a response and female subjects were faster when *self* and *female* shared a response. These correlations were strong for both the standard identity IAT and the *female–male/self–(other)* variant, but were weaker for the *male–female/other–(self)* variant

Table 3. Internal consistency and validity in Experiment 2 ($N = 67$)

| IAT or BIAT | Consistency and test–retest correlation | | | | Implicit–explicit correlation | | | | |
| | | | | | | Trial type | | Administration | |
	α	α_1	α_2	r_{12}	All	Focal	Non	1	2
Standard (seven-block) candidate attitude	.90	.87	.85	.62	.65	–	–	.62	.55
Kerry–Bush/good–(bad)	.77	.80	.71	.32	.63	.60	.59	.46	.58
Bush–Kerry/bad–(good)	.66	.46	.64	.26	.24	.22	.23	.17	.22
Standard (seven-block) gender identity	.93	.89	.89	.71	.76	–	–	.74	.67
Female–male/self–(other)	.92	.85	.84	.68	.67	.67	.57	.68	.58
Male–female/other–(self)	.67	.70	.50	.35	.46	.43	.38	.40	.35

Note. Internal consistencies α_1 and α_2 are for the two 64-trial BIAT measures that are intercorrelated (r_{12}) and for their combination (α). For the standard IAT, r_{12} is for the correlation between D measures based on Blocks 3 and 6 and ones based on Blocks 4 and 7 (see text). Validity is the correlation between corresponding implicit and explicit measures. Validity correlations are presented overall (All), across focal and nonfocal trials (focal and non), and administrations (1 and 2). For $N = 67$, r values associated with two-tailed α levels of .05 and .01 are .24 and .31, respectively.

(see Table 3). Internal consistencies were higher for both the standard identity IAT and the *self*-focal BIAT than for the *other*-focal BIAT.

For the standard identity IAT, the effect of subject gender emerged in the second step of the regression, $t(62) = 9.4$, $p = 10^{-13}$, zero-order $r = .76$, as it did for the *female–male/self–(other)* BIAT, $t(62) = 6.9$, $p = 10^{-8}$, zero-order $r = .67$. The effect of gender in the second step was also significant, but weaker, for the *male–female/other–(self)* BIAT, $t(62) = 3.6$, $p = .001$, zero-order $r = .46$. The standard IAT correlated .68 ($p = 10^{-9}$) with the *self*-focal gender-identity BIAT and .43 ($p = .0003$) with the *other*-focal gender-identity BIAT. The correlation between the two BIAT variants was $r = .56$, $p = 10^{-6}$.

Discussion

As in Experiment 1, the choice of focal attribute categories affected psychometric properties of BIAT measures. The BIATs that used *good* and *self* as focal categories showed strong convergence with their corresponding standard IATs. Compared to Experiment 1, the *bad*-focal and *other*-focal BIATs had slightly improved properties but were nevertheless inferior to the BIAT variants that that used *good* and *self* as focal categories. In combination, Experiments 1 and 2 strongly suggest that the BIAT method is suitable for attitude measurement when *good* is a focal category (and, apparently, only when *good* is a focal category) and for identity measurement when *self* is a focal category.

Experiment 3

Overview

Experiment 3 tested a shorter version of the BIAT. It also added an unanalyzed four-trial preface to each of its two trial blocks. These preliminary trials presented exemplars of the two concepts that switched responses between blocks twice each (e.g., *Bush* and *Kerry*). Exemplars of the two remaining

categories appeared on odd-numbered trials starting with Trial 5. The preliminary four trials were intended to assure that key assignments for the categories that would switch positions were effectively established prior to collection of data from the trials (Trials 5–20) to be used for computing D measures. Four new measurement topics were added to the two investigated in Experiments 1 and 2. The total of six BIATs included two attitude measures, two identity measures, and two stereotype measures.

Design

After a practice 20-trial block using nonsocial categories, all subjects completed a set of six BIATs. Each consisted of two 20-trial blocks and each was presented twice during the session. For half the subjects, the two identical BIATs appeared in immediate succession. For the remainder, the two repetitions were spaced so that the other five BIAT measures intervened. Half the subjects received the same category exemplars in both administrations and half received different exemplars (for cola brands and political candidates, the variety of available exemplars was limited and, therefore, exemplars were reused). The two attitude BIATs were *Kerry–Bush/good–(bad)* and *Coke–Pepsi/pleasant–(unpleasant);* the two identity BIATs were *female–male/self–(other)* and *Asian–American/self–(other)*; the two stereotype BIATs were *male–female/science–(arts)* and *African American–European American/weapons–(gadgets)*. The order of tasks within each BIAT was not varied, such that {*good+Kerry*}, {*pleasant+Coke*}, {*self+male*}, {*self+Asian*}, {*science+male*}, and {*weapons+African American*} always appeared before the complementary block.

Materials

Explicit Measures

Subjects answered three standard-format questions on 10-point scales for each of the six topics. The first two

requested judgments of the degree to which each concept category was associated with the focal attribute category. The third item was a measure of relative association in which the two concept categories were used as scale anchors. Explicit measures were scored by taking the difference between the ratings for the first two items and then weighting that equally with the third item. The one exception to this standard self-report format was gender identity. Subjects rated themselves on masculinity and femininity on 10-point scales and also reported their gender. The rating scales and stimulus sets are available in an online supplement[3].

Subjects and Procedure

One hundred and forty-nine subjects (109 females and 40 males) participated. After the *small/bird–(mammal)* practice BIAT, each of the six BIATs was administered twice, with the two identity BIATs first, the two attitude BIATs next, and the two stereotype BIATs last. In both identity BIATs, *self* was focal. In the attitude BIATs either *good* or *pleasant* was focal. In the stereotype BIATs, *science* (rather than *arts*) and *weapons* (rather than *gadgets*) were focal. Prior to the first administration of each BIAT, subjects answered its three associated self-report items. Subjects did the tasks in one of eight counterbalanced sequences that varied spacing between BIAT repetitions, novelty of items across repetitions, and whether each measure was first administered relatively early or relatively late in the procedure. The experiment required approximately 30 min to complete.

Results

Over all topics, and confirming observations in Experiments 1 and 2, mean latencies were generally faster on focal trials (679 ms) than on nonfocal trials (749 ms), $t(66) = 20.2$, $p = 10^{-44}$, and were also more accurate (5.5% vs. 7.6% errors, $t = 9.1$, $p = 10^{-13}$). The implicit–explicit correlations that support validity of BIAT measures were substantial for the two attitude BIATs and the two identity BIATs (see Table 4). For the two stereotype measures, the implicit–explicit correlations were positive, but small, which is typical for standard IAT measures of stereotypes.

In the hierarchical regressions that tested for procedural influences on IAT *D* measures, *spacing* and *novelty* served as design factors that, together with their interaction, were entered in the first step. The explicit measure was entered in the second step and the third step added interactions of spacing and novelty with the explicit measure. The effect of the explicit measure at its entry in the second step was large for all attitude and identity IATs and also reached statistical significance in the gender-science stereotype IAT, but was absent for the weapons–race IAT. As in the preceding

experiments, these hierarchical regressions revealed no effects that qualified the basic findings displayed in Table 4.

Discussion

Using a BIAT procedure with only two blocks of 20 trials each, Experiment 3 extended the evidence from Experiments 1 and 2 for validity and psychometric soundness of attitude and identity BIAT measures. The measures also had acceptable internal consistency. The BIAT measures of attitude and identity showed expected strong correlations with parallel self-report measures. This validity evidence did not vary as a function of novelty (or lack thereof) of exemplars in the second administration. Spacing between repetitions of identical BIATs had no significant effects on test–retest or implicit–explicit correlations.

Internal consistency and test–retest reliability of BIAT stereotype measures were somewhat lower than those for the attitude and identity BIATs. The relatively low implicit–explicit correlations for the two stereotype BIATs are not problematic. As previously noted, this is the pattern found for with standard IAT measures of stereotype (see Hofmann et al., 2005; Nosek, 2005; Nosek et al., 2007). At the same time, the ability of the BIAT measures to detect the same implicit stereotypes found in previous studies with standard IAT measures was less than clearly established. Mean *D* scores for the weapons–race and gender-science BIATs differed from zero in the expected directions, but were relatively small, at .17 and .16, relative to standard IATs (e.g., mean *D*s = .37 and .37 in Nosek et al., 2007).

Experiment 4

Together, results of Experiments 1–3 appear to have clearly established the usefulness of BIATs for assessing implicit attitudes and identities – even when used in the shortened (40 trial) format of Experiment 3. Experiment 3's uncertain appraisal of the success of BIATs for stereotype measures led to Experiment 4, which focused entirely on BIAT stereotype measures. For Experiment 4, analyses were planned to address questions for stereotypes of the form (illustrated here for the *race–weapons* IAT): Do the two BIAT variants, *Black–White/weapons–(gadgets)* and *White–Black/gadgets–(weapons)* result in measures that (a) are internally consistent, (b) detect the modal stereotypic association with equal sensitivity, and (c) are positively correlated with each other?

Experiment 4 used two BIAT variations in each of four stereotype domains. The training task in Experiment 3 was promoted to a full component of Experiment 4. Associations between size and type of animal were measured by *bird–mammal/small–(large)* and *mammal–bird/large–(small)* BIAT variations. A *disability–age* stereotype was assessed with *young–old/able–(disabled)* and *old–young/disabled–(abled)* variations. Gender stereotypes pertaining to academic

---

Table 4. Internal consistency and validity in Experiment 3 ($N = 149$)

| BIAT | Consistency and test–retest correlation | | | | Implicit–explicit correlation | | | | |
| | | | | | All | Trial type | | Administration | |
	α	α_1	α_2	r_{12}		Focal	Non	1	2
Female–male/self–(other)	.85	.79	.73	.67	.74	.74	.65	.69	.66
Asian–American/self–(other)	.72	.76	.52	.39	.48	.48	.43	.49	.30
Kerry–Bush/good–(bad)	.83	.79	.72	.56	.55	.49	.50	.46	.53
Coke–Pepsi/pleasant–(unpleasant)	.78	.71	.74	.38	.57	.54	.53	.49	.46
African American–European American/weapons–(gadgets)	.60	.54	.52	.20	.04	.05	.02	.10	−.04
Male–female/science–(arts)	.68	.68	.51	.24	.24	.25	.17	.16	.22

Note. Internal consistencies are for the two 32-trial BIAT measures (α_1 and α_2) that are intercorrelated (r_{12}) and for their combination (α). Validity is the correlation between corresponding implicit and explicit measures. Validity correlations are presented overall (All), across focal and nonfocal trials (focal and non), and administrations (1 and 2). For $N = 149$, r values associated with two-tailed α levels of .05 and .01 are .16 and .21, respectively.

disciplines were measured with *male–female/math–(arts)* and *female–male/arts–(math)* variants, and a *race–weapons* stereotype contrasted *Black–White/weapons–(gadgets)* with *White–Black/gadgets–(weapons)* variants. All were scored so that positive *D* scores would reflect the expected modal stereotype. That is, responses to blocks in which the focal categories were {*small+bird*}, {*large+mammal*}, {*disabled+old*}, {*able+young*}, {*science+male*}, {*arts+female*}, {*weapons+African American*}, and {*gadgets+European American*} were expected to be faster than those in the complementary blocks for each of the eight BIAT variants.

Materials

The self-report measures for each topic followed the pattern used in Experiment 3.

Subjects and Procedure

Ninety undergraduates (59 females and 31 males) completed the two BIAT variants for each of four stereotype domains. Each of the eight BIATs was done twice in succession using the 20-trial block structure introduced in Experiment 3, for a total of sixteen 40-trial BIATs. The two BIAT variants within each stereotype domain were consistently spaced by interposing one BIAT variant from each of the three other domains. Half of the subjects did all of the BIATs so that the task embodying the expected stronger association (e.g., {*disabled+old*}) preceded that with the alternative combination (e.g., {*disabled+young*}). Subjects were administered one of four task sequences that also counterbalanced the order of the two variations of each stereotype BIAT.

Results and Discussion

Consistent with observations of Experiments 1–3, across all eight BIATs, focal trials had shorter latencies than nonfocal trials (736 ms vs. 807 ms, $t = 12.1$, $p = 10^{-19}$) and fewer errors (8.5% vs. 11.8%, $t = 6.9$, $p = 10^{-9}$). Table 5 presents mean

IAT *D* scores, along with Cohen's *d* and internal consistencies for the eight BIATs. As in previous experiments, hierarchical regressions involving counterbalanced procedure variations did not qualify the findings presented in Table 5.

The findings in Table 5 show that all eight BIAT variants successfully detected the expected modal stereotypic association, with observed effect sizes ranging from moderate to large. Cohen's *d* values, which are presented for all of the measures in Table 5, ranged from 0.43 for *gadgets/White–Black* to 1.8 for *small/bird–(mammal)*. Internal consistencies for the eight pairs of BIATs (combining the two repetitions of each one) ranged from .63 to .78 and correlations between the two administrations of the same BIAT variant were positive, averaging $r = .29$. Only one set of stereotype BIATs displayed significant correlations with its set of parallel self-report measures. These were the two that associated gender with academic domains (rs = .38 and .38, $p \leq .0003$). All other implicit–explicit correlations were numerically positive, but none differed significantly from zero.

Not shown in Table 5 are correlations between the two variations (which differed in focal categories) of each stereotype BIAT. Positive correlations of at least moderate strength would suggest that the two variations were capturing the same stereotypic associations. This was found for two of the four domains: *Size–animal* ($r = .28$, $p = .008$) and *disability–age* ($r = .36$, $p = .001$). In those two domains, the corresponding explicit measures were also positively correlated with each other (rs = .44 and .55, respectively, ps $\leq .00002$). In the other two domains, the correlations were weaker: *Gender–academics* ($r = .14$, $p = .18$) and *race–weapons* ($r = .13$, $p = .24$). In these latter two domains, the corresponding explicit measures were also weakly intercorrelated (rs = .12 and .11, respectively, ps $\geq .25$).

General Discussion

The main features that distinguish the BIAT from standard IAT measures are (a) substantially fewer trials and (b) a task

Table 5. Effect size, internal consistency, and validity in Experiment 4 (*N* = 90)

Stereotype BIAT variant	Mean IAT *D* (Cohen's *d*)	Consistency and test–retest correlation				Implicit–explicit correlation
		α	α_1	α_2	r_{12}	
Bird–mammal/small–(large)	.71 (1.80)	.78	.77	.65	.31	.16
Mammal–bird/large–(small)	.65 (1.68)	.75	.77	.50	.17	.12
Young–old/able–(disabled)	.55 (1.62)	.63	.66	.45	.27	.06
Old–young/disabled–(able)	.58 (1.63)	.72	.64	.63	.35	.11
Male–female/math–(arts)	.33 (0.78)	.74	.65	.68	.33	.38
Female–male/arts–(math)	.37 (1.02)	.74	.71	.56	.17	.38
Black–White/weapons–(gadgets)	.30 (0.73)	.74	.66	.62	.32	.18
White–Black/gadgets–(weapons)	.17 (0.43)	.68	.61	.38	.36	.06

Note. All BIAT measures were scored so that societally modal stereotypes would receive numerically positive scores. *D* is the IAT score developed by Greenwald et al. (2003). Cohen's *d*, in parentheses, is the mean *D* divided by its standard deviation. Internal consistencies are for the two 32-trial BIAT measures (α_1 and α_2) that are intercorrelated (r_{12}) and for their combination (α). For *N* = 90, *r* values associated with two-tailed α levels of .05 and .01 are .21 and .27, respectively.

instruction to focus on just two of the four categories in each four-category test block. Subjects achieved the desired focus, as indicated by responses to the two focal categories being consistently faster and more accurate than responses to nonfocal categories.

In previous research, the attitude and identity topics of Experiments 1–3 had produced substantial positive correlations between standard IAT measures and parallel self-report measures. Explanation of this use of expected convergence of IAT and self-report measures to establish the utility of new IAT procedures appears in detail in Greenwald et al. (2003), pp. 199–200, 212). The BIAT measures of Experiments 1–3 indeed produced these positive correlations, although at slightly smaller magnitudes than for standard IAT measures, indicating that the BIAT functioned similarly to standard IAT measures. Experiment 1 also found two important exceptions to these large correlations and Experiment 2 replicated these exceptions. Specifically, the expected correlations with parallel self-report measures occurred strongly only when positive valence was focal for attitude BIATs and when *self* was focal for identity BIATs. They did not appear when negative valence was focal for an attitude BIAT or when *other* was focal for an identity BIAT.

For implicit stereotype measures, strong positive correlations with parallel self-report measures are not expected. Therefore, evidence for usefulness of stereotype BIATs was limited to observing whether BIAT results effectively detected several implicit stereotypes that were known, from previous research with standard IAT measures, to be observed pervasively (cf. Nosek et al., 2007). Experiment 4 confirmed that the BIAT method effectively detected these implicit stereotypes associated with age, race, and gender.

Limited Influence of Procedural Variables

Across Experiments 1–4, counterbalanced procedural factors included administration (first or second presentation of a specific BIAT), order of blocks within BIATs, spacing

between BIAT administrations, and the novelty of stimuli in the second presentation. Other than the "order effect", these procedural factors had no more than minor or inconsistent influences on BIAT measures.

In standard IAT measures, the associations tested first in sequence tend to appear stronger than those tested second. This was first reported by Greenwald et al. (1998) and was subsequently reported in numerous studies (e.g., Klauer & Mierke, 2005). This "order effect" has been speculatively attributed to associations being strengthened during performance of the first of the IAT's two combined tasks. Weak-to-moderate order effects for the political attitude BIAT in Experiment 1 and for two of the stereotype BIATs in Experiment 4 were observed. However, these effects were reversed in direction relative to order effects observed in standard IAT measures – that is, the associations assessed in the first combined task appeared weaker (not stronger) than those assessed in the second combined task. This finding may be a consequence of the limited practice preceding BIAT measures. This observation should be examined further in subsequent research.

Comparisons of present findings with previous findings, as well as the direct comparisons of BIATs and standard IATs built into present Experiment 2, indicated that the BIAT's reduction in trials produced relatively small decrements in psychometric performance on test–retest and implicit–explicit correlations. Such a result was perhaps anticipated by Brendl, Markman, and Messner (2001), who proposed that the IAT effect emerges, not at the level of single items, but at the level of complete test blocks.

Privileged Categories?

When *good* or *pleasant* was focal, attitude BIATs in the present research produced findings similar to those obtained previously with standard attitude IAT measures. Similarly, identity BIATs for which *self* was a focal category produced findings similar to those observed in previous IAT research. These observations suggested that, compared to their

complements, positive valence and *self* are, in some sense, privileged categories.

There was no corresponding evidence for privileged categories in the BIAT results for implicit stereotypes in Experiments 3 and 4. Nevertheless, for two of the four stereotype topics of Experiment 4 (gender–academics and race–weapons) variation of the BIAT's focal categories revealed asymmetries associated with the choice of focal categories.

Theoretical Interpretations

Associative Focus

Why do properties of BIAT measures vary with choice of focal categories? A possible theoretical explanation follows directly from the BIAT instructions which may induce an *associative focus* that allows the subject's performance to be determined primarily by a single association. For example, when asked to focus on *Pepsi* and *good*, subjects' associations of *Pepsi* with positive valence may become more accessible than other associations – viz., *Pepsi* with negative valence and *Coke* with either positive or negative valence. Several other researchers have aimed to achieve something resembling what we describe here as associative focus by designing IAT-like procedures that are limited to three categories – especially, the Go/No-go Association Test (Nosek & Banaji, 2001) and the Single Target and Single Category Association Tests (Bluemke & Friese, 2008; Karpinski & Steinman, 2006).

Valence Asymmetry

To explain the observed superior psychometric properties of attitude BIATs that use positive valence as a focal category requires not only the associative focus hypothesis, but also an assumption that the attitudes being measured in the present experiments are mentally represented more by positive than negative associations. The proposition that positive valence is cognitively more prominent than negative valence derives (in the modern era) from Zajonc's (1968) article on "attitudinal effects of mere exposure", which documented the greater frequency of positive than negative valence in various contexts, including lexicons. Support for this interpretation has recently appeared in the proposal by Unkelbach, Fiedler, Bayer, Stegmüller, and Danner (2008) that, compared to negatively valenced knowledge, positively valenced knowledge is more densely structured in memory. Unkelbach et al.'s density hypothesis implies that associative structures involving positive valence should be both more prominent and more cohesively structured than those involving negative valence (cf. Ashby, Isen, & Turken, 1999).

Self–Other Asymmetry

There is no theory of self–other asymmetry parallel to Unkelbach et al.'s (2008) density hypothesis for positive and negative valence. Nevertheless, such an assumption is quite plausible, in light of several scholarly treatments of the self that have described the self as drawing on memory structures that are considerably more complex than those that represent other persons (e.g., Greenwald, 1981; Kihlstrom & Cantor, 1984; Koffka, 1935).

Stereotype Asymmetry

In Experiment 4, correlations between pairs of stereotype BIATs that involve naturally complementary pairs (i.e., *able* complementary to *disabled*; *large* complementary to *small*) were relatively high (age–ability, $r = .36$; animals–size, $r = .28$). The two stereotype BIATs that showed weaker between-variant correlations were for gender–academics ($r = .14$) and race–weapons ($r = .13$). For these two BIATs, the associated category pairs (*math–arts* and *weapons–gadgets*) appeared to be less naturally complementary. They were also nominal categories, rather than adjectives.

Conclusions

The present findings imply that BIAT measures of implicit attitudes should have a positive focal valence category and BIAT measures of implicit identities should use *self* as a focal category. Stereotype measures based on the IAT may have distinct BIAT variants that differ as a function of the category chosen as focal.

We close by stating some questions that await resolution. First, can spaced repetitions of a BIAT increase test–retest reliability relative to the same number of trials used in a standard IAT format? Second, can multiple objects be placed on a common evaluative scale by making pairwise comparisons in individual attitude BIATs? Third, is it possible to use the BIAT procedure to measure attitudes on a scale that has a neutral-valence zero point? And, finally, are there coherent BIAT measures of implicit attitude that emerge from focusing on negative rather than positive valence?

Acknowledgements

age 11, was the chief tester during the development of the BIAT. The final version of the BIAT was the one among a dozen or more variations that reliably identified her as female in the gender identity task.

References

Aidman, E. V., & Carroll, S. M. (2003). Implicit individual differences: Relationships between implicit self-esteem, gender identity, and gender attitudes. *European Journal of Personality, 17*, 19–37.

Ashby, F. G., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review, 106*, 529–550.

Bluemke, M., & Friese, M. (2008). Reliability and validity of the Single-Target IAT (STIAT): Assessing automatic affect towards multiple attitude objects. *European Journal of Social Psychology, 38*, 977–997.

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

Farnham, S. D., & Greenwald, A. G. (1999, June). *In-group favoritism = implicit self-esteem X in-group identification.* Paper presented at meetings of the American Psychological Society, Denver, CO.

Greenwald, A. G. (1981). Self and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (*Vol. 15, pp. 201–236)*. New York: Academic Press.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*, 3–25.

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

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*, 1369–1385.

Inquisit 2.0 [Computer software]. (2005). Seattle, WA: Millisecond Software LLC.

Karpinski, A., & Steinman, R. B. (2006). The Single Category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*, 16–32.

Kihlstrom, J. F., & Cantor, N. (1984). Mental representations of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (*Vol. 17, pp. 2–48)*. New York: Academic Press.

Klauer, K. C., & Mierke, J. (2005). Task-set inertia, attitude accessibility, and compatibility-order effects: New evidence for a task-set switching account of the Implicit Association Test effect. *Personality and Social Psychology Bulletin, 31*, 208–217.

Koffka, K. (1935). *Principles of gestalt psychology*. New York: Harcourt.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*, 565–584.

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-go Association Task. *Social Cognition, 19*, 625–664.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., et al. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*, 36–88.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 3*, 416–442.

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

Rudman, L. A., Greenwald, A. G., & McGhee, D. E. (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin, 27*, 1164–1178.

Sriram, N., Nosek, B. A., & Greenwald, A. G. (2007). *Scale invariant contrasts of response latency distributions.* Unpublished manuscript, University of Virginia.

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology, 95*, 36–49.

West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality, 64*, 1–48.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1–27.

Anthony G. Greenwald

Department of Psychology
University of Washington
Seattle
WA 98195
USA
Tel. +206 543 7227
Fax +206 685 3157
E-mail agg@u.washington.edu