The Brief Implicit Association Test

N. Sriram
University of Virginia

Anthony G. Greenwald
University of Washington

N. Sriram
Department of Psychology
University of Virginia
Charlottesville, VA 22904

nsriram@virginia.edu

corresponding author
Anthony G. Greenwald
Department of Psychology
University of Washington
Seattle, WA 98195

agg@u.washington.edu
voice: (206) 543-7227
FAX: (206) 685-3157

Abstract

The Brief Implicit Association Test (BIAT) consists of 2 blocks of trials (total: 40–64 trials) that use the same 4 categories and the same stimulus-response mappings as the 2 combined tasks of a standard IAT (176 trials). The BIAT's instructions focus the subject's attention on just 2 of the 4 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 6 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.

Word count = 149

In ten 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 vs. negative valence (e.g., *pleasant* vs. *unpleasant*). Identity (or self-concept) IATs combine contrast of *self vs. 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 task of Nosek & Banaji, 2001 and the single-category IATs of Karpinski & Steinman, 2006 and Wigboldus, in press) 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 hoped to reduce spontaneous variation in subject strategy.

The IAT has four categories, each mapped onto one of two responses. The BIAT instructs respondents to focus on just two of the IAT's category–response mappings in each combined task. One category is focal in both combined tasks, meaning that one other category is focal in neither. Prior to each of the two tasks, 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 the "match" (or "yes") response key, and (c) to respond to any other stimuli with an alternative "mismatch" (or "no") response key. With two blocks and a total of fewer than 80 trials, the BIAT substantially reduces administration time relative to the standard 5-step IAT procedure, which is often done with seven blocks of trials, which usually involve approximately 180 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

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

Design of the Brief IAT

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). All BIATs use exemplars of four categories, but only three are focal during the two combined tasks. The category that is focal in both combined tasks (*pleasant* in the example just given) has a contrasting category that remains non-focal in both tasks (*unpleasant* in this example). BIAT names list the four categories, placing the category that remains non-focal last and marking it also with 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.

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). These 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 seconds to process the BIAT block instructions.

After the instruction page display, the lists of exemplars of the focal categories 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 the correct response. This created a built-in error penalty, which is also a property of standard IAT measures (cf. Greenwald, et al., 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 2.0 (Millisecond Inc., 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, Nosek, & Banaji, 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).

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 16-trial blocks of Experiments 3 and 4, one subset had trials {1, 2, 5, 6, 11, 12, 15, 16} and the other subset had the remaining eight trials. *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 also of a *test–retest* estimate of reliability.

Explicit attitude measures include (a) the difference between separately rated strengths of association of contrasted concepts with positive or negative valence (e.g., difference between liking ratings for cola brands) and (b) single-item Likert-format measures of relative preference between the contrasted categories. 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 societally pervasive in previous research (Nosek et al., 2007b). 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 step of these 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 explicit 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. An attitude BIAT contrasting valence associations with the two candidates assessed implicit candidate preference. An identity BIAT, measuring the association between self and gender was also included. 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, should 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)*. Applying the previously introduced naming convention, 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 *Bush* 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 two repetitions in immediate succession of each of the four 64-trial BIATs, including both variants of the candidate attitude and the gender identity measures. 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 and the 2–block BIAT sequence was repeated in succession. 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, 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, 10=very warm) towards each of George Bush, John Kerry, the Republican Party and the Democratic Party, and then reported political identity along a continuum anchored by Democrat and Republican. Next, subjects practiced the BIAT instructions using two non-social focal categories — *curved* (circle, oval, ring, ball) and *bird* (eagle, swan, parrot, duck) — in a single 32-trial training block. Non-focal 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 order of combined blocks within each BIAT, the placement of the attitude and identity measures in the first or second half of the sequence, and the position of the two BIAT variants within each half.

Results

Correct responses to focal category items can be called *match* responses; those to non-focal categories are *mismatch* responses, Across the four BIATs, match responses were faster (709 ms) than mismatch responses (774 ms), $t(39) = 9.58$, $p = 10^{-11}$ and also had fewer errors, 8.5% vs. 10%, $t(39) = 2.85$, $p = .007$. These findings were consistent with the expectation that subjects would adopt the instructed strategy of focusing on the instructed 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 self-reported candidate attitude and the average of the two *Kerry–Bush/good–(bad)* BIATs, was .761 (see Table 1).

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, $t(35) = 2.18$, $p = .04$). The expected strong effect of explicit (self-report) preference emerged clearly in the second step, $t(35) = 6.62$, $p = 10^{-7}$, zero-order $r = .761$. 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 {*baḍ*+*Kerry*} and {*baḍ*+*Bush*} blocks were 712 ms and 778 ms for Bush supporters, compared to 842 ms and 810 ms for Kerry supporters. Unexpectedly, neither the 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)* BIAT were not significantly intercorrelated, $r = .175$, $p = .28$.

Gender identity BIATs. In the *female–male/self–(other)* BIAT dichotomous subject gender served as the explicit measure and correlated $r = .700$ 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 $r = .551$, $p = .0003$. In the hierarchical regression, the expected main effect of the self-reported sex was evident in the second step, $t(35) = 5.06$, $p = 10^{-5}$, zero-order $r = .700$, 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 ms and 750 ms for male subjects and were 715 ms and 690 ms for female subjects. Contrary to expectations, for this BIAT, both internal consistency and test–retest correlation for the *D* measure were low (see Table 1). The *D* measure was uncorrelated with subject gender and was also unrelated to the *female–male/self–(other)* variant, $r = -.035$. 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 (*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 non-salient 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 non-figural 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 non-salient) 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 7-block IAT measures. Subjects provided data for both formats.

Design

The standard IAT is a 5-step, 7-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 one on Blocks 4 and 7 (Greenwald et al., 2003). The correlation between the measure based on the contrast of Blocks 3 and 6 and the one based on the contrast of Blocks 4 and 7 provided one measure of internal consistency. Using the procedure in the General Method section, a second internal consistency measure was computed for the two contrasts separately. 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", 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 match and mismatch 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 match than mismatch trials (726 ms vs. 777 ms, $t(66) = 7.57$, $p = 10^{-9}$) and were also more accurate (7.8% vs. 8.7% errors, $t(66) = 2.48$, $p = .016$), indicating adherence to task instructions. 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 (22 males, 45 females) 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 2). 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, $t(62) = 6.47$, $p = 10^{-7}$, zero-order $r = .647$, and the *Kerry–Bush/good–(bad)*BIAT, $t(62) = 6.75$, $r = .651$, $p = 10^{-8}$, zero-order $r = .632$, Like Experiment 1, the effect of explicit preference was not significant in the hierarchical regression of the *Bush–Kerry/bad–(good)* BIAT, $t(62) = 1.56$, $r = .222$, $p = .13$. The standard IAT correlated $r = .649$ ($p = 10^{-8}$) with the *good*-focal BIAT, but only $r = .389$ ($p = .001$) with the *bad*-focal BIAT. The correlation between the two BIAT variants was $r = .289$, $p = .02$.

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 (see Table 2). Internal consistencies were higher for both the standard identity IAT and the *self*-focal BIAT than for the *other*-focal BIAT.

In regression analysis of the standard identity IAT, the effect of subject gender emerged in the second step of the regression, $t(62) = 9.43$, $p = 10^{-13}$, zero-order $r = ,762$, as it did for the *female–male/self–(other)* BIAT, $t(62) = 6.89$, $p = 10^{-8}$, zero-order $r = ,665$. The effect of gender in the second step was also significant, but weaker, for the *male–female/other–(self)* BIAT, $t(62) = 3.58$, $p = .001$, zero-order $r = ,459$. The standard IAT correlated .680 ($p = 10^{-9}$) with the *self*-focal gender-identity BIAT and .434 ($p = .0003$) with the *other*-focal gender-identity BIAT. The correlation between the two BIAT variants was $r = .561$, $p = 10^{-6}$.

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 than used in Experiments 1 and 2. It also added an unanalyzed 4-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 categories for which assigned responses did not change 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 non-social 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 other BIAT measures intervened. Half the subjects received the same category exemplars in both administrations and half received different exemplars (excepting the tests for cola brands and political candidates, for which the available variety of available exemplars was quite limited). 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 constant in that {*good*+*Kerry*}, {*pleasant*+*Coke*}, {*self*+*male*}, {*self*+*Asian*}, {*science*+*male*}, and {*weapons*+*African American*} always appeared before the complementary block.

Materials

Explicit measures. With one exception, 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 that, for gender identity, subjects were asked to rate themselves on masculinity and femininity on 10-point scales and to report their gender (see Appendix for rating scales and stimulus sets).

BIAT stimuli. The initial practice BIAT used the categories *small* (tiny, little, small, light), *big* (huge, big, massive, heavy), *bird* (robin, sparrow, parrot, duck), and *mammal* (elephant, rhino, bison, giraffe). Subsequent BIATs (with the exception, explained above, of the cola and candidate tests) used two sets of four exemplars in each category to study the effect of stimulus novelty across administrations.

Subjects and Procedure

One hundred forty-nine subjects (109 females, 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 minutes to complete.

Results

Over all topics, and confirming observations in Experiments 1 and 2, mean latencies were generally faster on match trials (679 ms) than on mismatch trials (749 ms), $t(66) = 20.18$, $p = 10^{-44}$, and were also more accurate (5.5% vs. 7.6% errors, $t = 9.05$, $p = 10^{-13}$). The validity correlations between BIAT measures and the corresponding IAT *D* measures were substantial for the attitude and identity BIATs (see Table 3). Implicit–explicit correlations were substantial for both the attitude and identity BIATs (see Table 3). 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 3.

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 (see Table 3). 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., 2007b). 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., 2007b).

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 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 Appendix describes stimuli used for all tests, as well as the rating scales used for parallel self-report measures. The self-report measures for each topic were combined as for Experiment 3 to assess the explicit stereotype parallel to each BIAT measure.

Subjects and Procedure

Ninety undergraduates (59 females, 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, match trials had shorter latencies than mismatch trials (736 ms vs. 807 ms, $t = 12.07$, $p = 10^{-19}$) and fewer errors (8.5% vs. 11.8%, $t = 6.93$, $p = 10^{-9}$). Table 4 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 4.

The findings in Table 4 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 4, ranged from 0.43 for *gadgets/White–Black* to 1.80 for *small/bird–mammal*. Internal consistencies for the eight pairs of BIATs (combining the two repetitions of each one) ranged from .630 to .778 and correlations between the two administrations of the same BIAT variant were positive, averaging $r = .287$. Only set of stereotype BIATs displayed significant correlations with their parallel self-report measures. These were the two that associated gender with academic domains (rs = .377 and .375, $p \leq .0003$). All other implicit–explicit correlations were numerically positive, but none differed significantly from zero.

Not shown in Table 4 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 = .279, p = .008) and *disability–age* (r = .359, p = .001). In those two domains, the corresponding explicit measures were also positively correlated with each other (rs = .442 and .550, respectively, ps ≤ .00002). In the other two domains, the correlations were weaker: *gender–academics* (r = .143, p = .18) and *race–weapons* (r = .126, p = .24). In these latter two domains, the corresponding explicit measures were also weakly intercorrelated (rs = .123 and .110, respectively, ps ≥ .25).

General Discussion

The main features that distinguish the BIAT from standard IAT measures are (a) substantially fewer trials, and (b) a task instruction to focus on just two of the four categories in each 4-category test block. All four of the present experiments clearly confirmed that subjects achieved the desired focus on two categories. That is, responses to the two focal categories in each 4-category task were significantly faster and more accurate than those to the task's two non-focal categories.

The attitude and identity topics of Experiments 1–3 had been selected for the present research because of previous findings that these topics produced substantial positive correlations between standard IAT measures and parallel self-report measures. The reasoning underlying use of this expected convergence of IAT and self-report measures to establish usefulness of IAT measures has been described in detail by Greenwald et al. (2003, pp. 199–200, 212). Finding similarly substantial correlations between BIAT measures and parallel self-report measures would therefore indicate that the BIAT was functioning similarly to standard IAT measures. The BIAT

measures of Experiments 1–3 indeed produced these positive correlations, although at slightly smaller magnitudes than for 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 when positive valence was focal for attitude BIATs and when *self* was focal for identity BIATS. However, 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 characteristically observed. Therefore, evidence for usefulness of stereotype BIATs was limited to observing whether the BIAT method could effectively detect several implicit stereotypes that were known, from previous research with standard IAT measures, to be observed pervasively (cf. Nosek et al., 2007b). Experiment 4 confirmed that the BIAT method effectively detected these implicit stereotypes associated with age, race, and gender.

Privileged Categories?

When *good* or *pleasant* (i.e., positive valence) 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* (but not *other*) was a focal category produced findings similar to those observed in previous IAT research. These findings suggested that, compared to the complementary categories (negative valence and *other*) positive valence and *self* are in some sense privileged categories.

There was no corresponding evidence for privileged categories in the results for BIAT measures of 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 affected findings. These results again revealed asymmetries associated with choice of categories made focal in the BIAT procedure.

Theoretical Interpretation: Associative Focus

As just summarized, in comparing pairs of BIATs that used the same sets of four categories while varying which were focal in the two combined-task blocks, Experiments 1, 2, and 4 all found that properties of BIAT measures varied with choice of focal categories. A possible theoretical explanation follows directly from the key design feature of the BIAT in requesting focus, in each of the two combined-task blocks, on just two of the task's four categories. This instruction may induce an *associative focus* that allows the subject's performance to be determined primarily by a single association. For example, when subjects in Experiment 3 were asked to focus on *Pepsi* and *good*, their associations of *Pepsi* with positive valence may have become more accessible than were other associations involving the task's four categories — 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 Category Association Test (Karpinski & Steinman, 2006; Wigboldus, in press).

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, include lexicons. Support this interpretation has recently appeared in the proposal by Unkelbach, Fiedler, Bayer, Stegmuller, 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 composed of the same four categories showed relatively weak intercorrelations (averaging r = .233) between the two variations that had different focal categories. The two of these that showed higher correlations (category sets: *young, old; able, disabled*, r = .359; and *mammal,bird,large,small,* r = .279) appear to involve naturally complementary pairs (i.e., *able* complementary to *disabled*; *large* complementary to *small*). For the two stereotype BIATs that showed weaker correlations between their two versions (gender–academics, r = .143 and race–weapons, r = .126) the associated category pairs (*math–arts* and *weapons–gadgets*) differed both in appearing to be less complementary and in being nominal categories, rather than adjectives.

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. With only one exception — described in the next paragraph — 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. The present research found weak-to-moderate order effects for the political attitude BIAT in Experiment 1 and for two of the stereotype BIATs in Experiment 4. However, these effects were reversed in direction relative to those most likely to occur with 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, a possibility that should be examined in subsequent research.

Comparisons of present findings with previous findings, as well the direct comparisons of BIAT and standard IATS built into present Experiment 2, indicate that the BIAT's reduction in trials produced relatively small decrements in psychometric performance on test–retest and implicit–explicit correlations. Such performance of the BIAT 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.

Potential Efficiency of the BIAT

Figure 2 indicates the potential for repeated administrations of BIAT measures to increase test–retest reliability to levels that are conventionally deemed satisfactory for individual difference measures. Standard IAT measures have test–retest reliabilities that had a median of $r = .56$ across nine available reports (reviewed by Nosek, Greenwald, & Banaji, 2007a). Figure 2 applies the Spearman–Brown prediction formula to data from present Experiments 1, 2, and 3 to estimate expected test–retest reliability of repeated administrations of BIAT measures. Figure 2's estimated test–retest reliability of single 40-trial BIAT measures was averaged across Experiment 3's two attitude (political and soft-drink) and two identity (gender and Asian ethnicity) BIATs. Estimated test–retest reliabilities of 64-trial BIAT measures were also averaged across four tests, provided by the political attitude and gender identity BIATs of Experiments 1 and 2. Average test–retest reliabilities were nearly identical for 40-trial and 64-trial BIAT measures. For comparison, Figure 2 also includes test–retest reliabilities for standard IAT measures, using the estimate of $r = .56$ from Nosek et al. (2007a).

Conclusions from Figure 2 are necessarily limited by two factors: (a) the small number of data sets from which test–retest reliabilities were estimated and (b) the close temporal proximity of repeated administrations of BIATs. The latter factor may not be a major concern, given that test–retest reliabilities of IAT measures have been observed to be relatively independent of test–retest interval in previous research (Nosek et al., 2007a, Figure 6.1). Nevertheless, it is conceivable that the close temporal proximities of test and retest for BIAT measures in the present research has overestimated test–retest correlations that can be expected when they are more widely separated. The presently observed test–reliabilities of single BIAT measures were high enough to suggest that two repetitions of a 40-trial BIAT measure (80 trials) can provide

test–reliability exceeding that of a standard IAT measure that is often approximately 180 trials in length. Likewise (in theory), three repetitions of a 40-trial BIAT measure (120 trials) may provide test–retest reliabilities exceeding $r = .75$, a level that many consider sufficient to permit use in assessing individual differences.

Associative Focus Considered Further

The present associative focus theoretical interpretation, along with the hypothesized asymmetries of positive–negative valence and *self –other*, has two implications for how BIAT measures may be most effectively used in research. Specifically, BIAT measures of implicit attitude should have a positive (rather than negative) valence category focal and BIAT measures of implicit identities should use *self* (rather than *other*) as a focal category. Further investigations could usefully assess the generality of these conclusions. A few questions that seem especially worth pursuing are: How general is the observation that BIAT attitude measures are more valid when positive valence is focal? For example, might BIAT measures for objects of typically negative attitudes (such as war, pollution, and crime) work well when negative valence is focal? Similarly, how general is the observation that BIAT identity measures are more valid when *self* is focal?

We close by suggesting three other research questions that, if pursued, could extend understanding of the usefulness of BIAT measures in research and practice. First, can the implications of present Figure 2 can be confirmed? That is, can three repetitions of a BIAT measure be relied on to provide an implicit measure that has test–retest reliability exceeding $r = .75$ when test and retest are temporally separated. Second, can attitude BIATs be designed to measure attitudes on a scale that has a neutral-valence zero point? For example, would a *good–bad/Nixon–(Kennedy)* BIAT allow determination of whether the respondent is attitudinally

positive versus negative to *Richard Nixon*? And third, what is the role of the BIAT category that remains non-focal? The associative focus interpretation implies that it is of minor importance. However (for example), might the result of a *good–bad/Nixon–(Kennedy)* BIAT be different from that of *good–bad/Nixon–(Reagan)*?

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.

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.

Egloff, B., & Schmukle, S. C. (2002). Predictive validity of an implicit association test for assessing anxiety. *Journal of Personality and Social Psychology*, *83*, 1441-1455.

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., & 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., & Klauer, K. C. (2005). Validity of the Salience Asymmetry interpretation of the Implicit Association Test: Comment on Rothermund and Wentura. *Journal of Experimental Psychology: General*, *134*, 420–425.

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

Greenwald, A. G., Poehlman, T. A., E. L. Uhlmann, & Banaji, M. R. (in press). Understanding and using the Implicit Association Test: III. Meta-analysis of Predictive Validity. *Journal of Personality and Social Psychology*.

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.

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. *Principles of gestalt psychology*. New York: Harcourt, 1935.

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., Greenwald, A. G., & Banaji, M. R. (2007a). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes* (pp. 265–292). Psychology Press.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007b). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*, 36–88.

Perugini, M. (2005). Predictive models of implicit and explicit attitudes. *British Journal of Social Psychology*, *44*, 29-45.

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

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.

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

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.

Wigboldus, D. (in press). Single target implicit associations. *Experimental Psychology*.

Winer, B. J. (1962). *Statistical principles in experimental design*. New York: McGraw–Hill.

Unkelbach, C., Fiedler, K., Bayer, M., Stegmuller, M., & Danner, D. (in press). Why postive information is processed faster: The Density hypothesis. *Journal of Personality and Social Psychology*.

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

Acknowledgment Note

The University of Washington has applied for patent on the BIAT method. The patent is managed by Project Implicit, a non-profit organization of which the second author is an officer. Both the University of Washington and Project Implicit authorize free use of the BIAT method and published stimuli for scholarly research, provided that reports of the research clearly identify any modifications made to the BIAT and appropriately cite the present article. Please contact Project Implicit (feedback@projectimplicit.net) to request a license for commercial or other non-scholarly use of the BIAT. The authors thank Justin Storbeck and Yoav Bar-Anan for comments on earlier drafts.

Footnotes

1. Variation in this interval (150 ms to 750 ms) was previously shown to have no significant impact on IAT measures (Greenwald et al., 1998).
2. Several researchers have indicated reluctance to counterbalance orders of combined-task blocks in IAT measures, assuming that the order effect has potential to weaken correlational findings in studies that use IAT measures to assess individual differences (e.g., Egloff & Schmukle, 2002; Perugini, 2005). In the previous experience of the authors, which was confirmed again in the present research, order effects typically play no more than small, and typically negligible, roles in moderating correlation magnitudes .

Appendix

Explicit Measures in Experiment 3

Gender Identity	Does the trait Masculine describe you? Does the trait Feminine describe you? Indicate whether you are Male or Female	1 = Not at all 1 = Not at all 1 = Male	10 = Very Well 10 = Very Well 2 = Female
Ethnic Identity	How much do you identify with American culture? How much do you identify with Asian culture? Which culture do you identify with more?	1 = Not at all American 1 = Not at all Asian 1 = Strongly Asian	10 = Strongly American 10 = Strongly Asian 10 = Strongly American
Candidate Attitude	Rate how you feel about George Bush. Rate how you feel about John Kerry. Who do you prefer, George Bush or John Kerry?	1 = Very Cold 1 = Very Cold 1 = Strongly Prefer Bush	10 = Very Warm 10 = Very Warm 10 = Strongly Prefer Kerry
Cola Attitude	How much do you like Coca Cola? How much do you like Pepsi Cola? Which do you prefer, Coke or Pepsi?	1 = Dislike Coke 1 = Dislike Pepsi 1 = Prefer Coke	10 = Like Coke 10 = Like Pepsi 10 = Prefer Pepsi
Weapons Stereotype	How strongly do you associate Weapons and African Americans? How strongly do you associate Weapons and European Americans? Which group has stronger associations with weapons?	1 = Weak Association 1 = Weak Association 1 = African Americans	10 = Strong Association 10 = Strong Association 10 = European Americans
Science Stereotype	How strongly do you associate Science with Females? How strongly do you associate Science with Males? Which group has stronger associations with Science?	1 = Weak Association 1 = Weak Association 1 = Females	10 = Strong Association 10 = Strong Association 10 = Males

IAT Word Exemplars in Experiment 3

self	I, self, me, myself	my, self, mine, I
other	other, them, they, their	other, they, it, theirs
male	man, male, he, brother	boy, male, him, son
female	woman, female, she, sister	girl, female, her, daughter
Asian	Curry, Karate, Beijing, Sony	Noodles, Taekwondo, Shanghai, Honda
American	Burger, NFL, Boston, Microsoft	Hot Dog, NBA, Houston, Chrysler
good	freedom, peace, joy, success	love, smile, friend, honest
bad	abuse, poison, failure, enemy	evil, frown, ugly, sick
pleasant	nice, pleasure, gold, happy	Great, heaven, lucky, sunshine
unpleasant	nasty, hell, pollution, vomit	horrible, agony, stink, rotten
science	engineering, chemistry, laboratory, molecule	physics, experiment, technology, equation
arts	literature, french, poetry, music	history, writing, spanish, painting
male	John, Paul, Mike, Kevin	Steve, Greg, Jeff, Bill
female	Amy, Joan, Lisa, Sarah	Diana, Kate, Ann, Rachel

Explicit Measures in Experiment 4

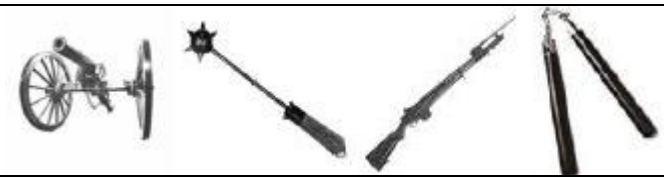
size/animal	Rate the size of members of the category birds. Rate the size of members of the category mammals. Which, in your opinion, are larger, birds or mammals?	1 = Very small 1 = Very small 1 = Birds are larger	9 = Very large 9 = Very large 9 = Mammals are larger
Able/ old−young	How much do you associate being able bodied with being young? How much do you associate being able bodied with being old? Which group has stronger association with able, young or old?	1 = Not at all 1 = Not at all 1 = Young	9 = Very Strongly 9 = Very Strongly 9 = Old
Disabled/ young−old	How much do you associate being disabled with being young? How much do you associate being disabled with being old? Which group has a stronger association with disabled, young or old?	1 = Not at all 1 = Not at all 1 = Young	9 = Very Strongly 9 = Very Strongly 9 = Old
Math/ female−male	How much do you associate math with male? How strongly do you associate math with female? Do males or females have a stronger association with math?	1 = Not at all 1 = Not at all 1 = Male	9 = Very Strongly 9 = Very Strongly 9 = Female
Arts/ male−female	How much do you associate arts with male? How strongly do you associate arts with female? Do males or females have a stronger association with arts?	1 = Not at all 1 = Not at all 1 = Male	9 = Very Strongly 9 = Very Strongly 9 = Female
Weapons/ white−black	How strongly do you associate weapons and African Americans? How strongly do you associate weapons and European Americans? Which group has stronger associations with weapons?	1 = Weak Association 1 = Weak Association 1 = African Americans	9 = Strong Association 9 = Strong Association 9 = European Americans
Gadgets/ black−white	How strongly do you associate gadgets and African Americans? How strongly do you associate gadgets and European Americans? Which group has stronger associations with gadgets?	1 = Weak Association 1 = Weak Association 1 = African Americans	9 = Strong Association 9 = Strong Association 9 = European Americans

IAT Word Exemplars in Experiment 4

small	tiny, little, small, light
large	big, massive, huge, heavy
mammal	elephant, bison, rhino, giraffe
bird	robin, sparrow, parrot, duck
math	math, algebra, calculus, equation
arts	poetry, dance, literature, art
male	man, male, boy, brother
female	woman, female, girl, sister

Experiments	Category	Concept Exemplars used in Attitude IATs
1, 2, 3	George Bush	
1, 2, 3	John Kerry	
3	Coke	
3	Pepsi	

Experiment	Category	Race Exemplars used in Stereotype IATs
3	African American	
3	African American	
3	European American	
3	European American	
4	African American	
4	European American	

Category	Object Exemplars used in Stereotype IATs (Experiments 3 and 4)
Weapons	
Weapons	
Gadgets	
Gadgets	

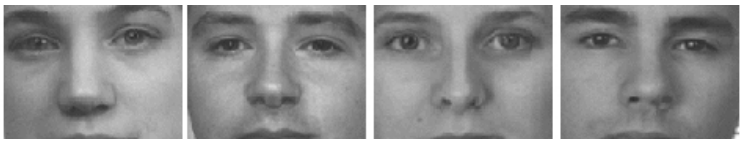
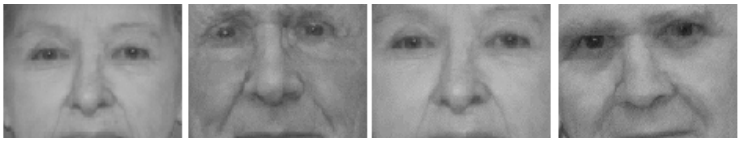
Category	Exemplars used in Disability–Age Stereotype IATs (Experiment 4)
Able	
Disabled	
Young	
Old	

TABLE 1

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}		Mat	Mis	1	2
Kerry–Bush/good–(bad)	.827	.815	.659	.489	.761	.753	.646	.701	.601
Bush–Kerry/bad–(good)	.757	.785	.656	.167	.105	–.015	.218	.049	.116
female–male/self–(other)	.936	.924	.826	.551	.700	.728	.604	.734	.466
male–female/other–(self)	.551	.617	.506	.067	.068	–.143	.247	.059	.038

[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 match and mismatch trials (Mat and Mis), and for first or second BIAT administrations (1 and 2). For N = 40, *r* values associated with 2-tailed alpha levels of .05 and .01 are .304 and .393, respectively.

TABLE 2

Internal Consistency and Validity in Experiment 2 (N=67).

IAT or BIAT	Consistency and test–retest correlation				Implicit–Explicit Correlation				
					All	Trial Type		Administration	
	α	α_1	α_2	r_{12}		Mat	Mis	1	2
standard (7-block) candidate attitude	.900	.869	.846	.624	.647	–	–	.620	.547
Kerry–Bush/good–(bad)	.774	.799	.708	.320	.632	.599	.593	.459	.578
Bush–Kerry/bad–(good)	.656	.455	.638	.256	.244	.217	.230	.165	.220
standard (7-block) gender identity	.930	.889	.886	.709	.762	–	–	.739	.665
female–male/self–(other)	.915	.852	.835	.675	.665	.668	.566	.680	.578
male–female/other–(self)	.669	.696	.500	.345	.459	.428	.384	.404	.345

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 match and mismatch trials (Mat and Mis), and administration (1 and 2). For N = 67, *r* values associated with 2-tailed alpha levels of .05 and .01 are .250 and .325, respectively.

TABLE 3

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}		Mat	Mis	1	2
female–male/self–(other)	.847	.787	.728	.673	.739	.738	.652	.687	.664
Asian–American/self–(other)	.723	.759	.524	.385	.478	.476	.430	.487	.298
Kerry–Bush/good–(bad	.827	.793	.715	.564	.553	.486	.504	.457	.525
Coke–Pepsi/pleasant–(unpleasant)	.781	.707	.738	.382	.570	.543	.529	.485	.463
African American–European American/weapons–(gadgets)	.595	.539	.522	.201	.044	.053	.024	.100	–.035
male–female/science–(arts)	.679	.680	.509	.244	.240	.249	.172	.160	.222

Note: Internal consistencies are for the two 32-trial IAT measures (α_1 and α_2) that are inter–correlated (r_{12}) and for their combination (α). Validity is the correlation between corresponding implicit and explicit measures. Validity correlations are presented overall (All), across match and mismatch trials (Mat and Mis), and administration (1 and 2). For N = 149, *r* values associated with 2-tailed alpha levels of .05 and .01 are .160 and .210, respectively.

TABLE 4

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)	.708 (1.80)	.778	.766	.650	.313	.161
mammal-bird/large–(small)	.653 (1.68)	.747	.774	.499	.167	.119
young–old/able–(disabled)	.551 (1.62)	.630	.662	.454	.270	.056
old–young/disabled–(able)	.583 (1.63)	.718	.642	.630	.349	.108
male–female/math–(arts)	.330 (.783)	.738	.645	.667	.331	.375
female–male/arts–(math)	.372 (1.02)	.736	.712	.560	.176	.377
Black–White/weapons–(gadgets)	.302 (.728)	.744	.660	.622	.323	.180
White–Black/gadgets–(weapons)	.170 (.425)	.677	.613	.382	.355	.062

Note: All IAT measures were scored so that societally modal stereotypes would receive numerically positive scores. *D*. is the IAT score developed by Grenwald 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 IAT measures (α_1 and α_2) that are inter–correlated (r_{12}) and for their combination (α). For N = 90, *r* values associated with 2-tailed alpha levels of .05 and .01 are .205 and .267, respectively.

Figure Captions

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

Figure 2. Comparison of observed and projected test–retest correlations for standard (176-trial) IAT with 40-trial and 64-trial versions of BIAT. Estimates of test–retest correlations for 40-trial and 64-trial BIAT are averaged over four estimates obtained from Experiments 1–3. The estimate for the standard IAT is the median reported by Nosek et al. (2007a). Theoretical curves based on test–retest reliability $r = .45$ and $r = .65$ are shown for comparison.

{Pepsi}

{PLEASANT}

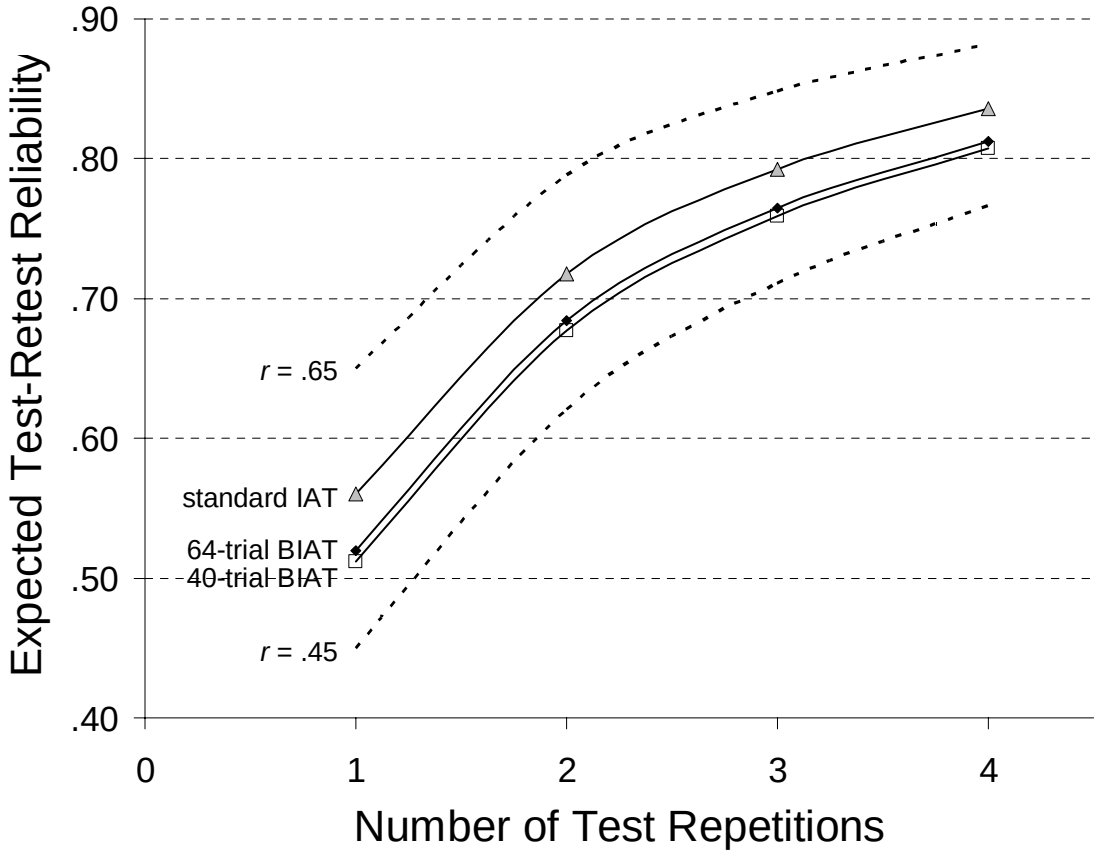
NICE, HEAVEN, HAPPY, PLEASURE

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.


.90
.80
.70
.60
.50
.40
Expected Test-Retest Reliability
r = .65
standard IAT
64-trial BIAT
40-trial BIAT
r = .45
0
1
2
3
4
Number of Test Repetitions