

We made the subjects respond with the right hand to the stimulus on the right side, and with the left hand to the stimulus on the left side. . . . When movement of the right hand was required with stimulation on the left side or the other way round, then the time lapse was longer and errors common. (Donders, 1868/1969, p. 421)

With this modest test Donders opened up possibilities for studying mental processes, the effects of which are visible in behavioral and brain science even 150 years later. The idea underlying an entire family of response latency techniques remains the same as conjectured and tested by Donders: the easier a mental task, the quicker the decision point is reached and the fewer the errors that result. To make the right–right and left–left association is easier than the right–left and left–right association, and the difference in speed between the two tasks can serve as an indicator of their relative difficulty. With a psychology that relied on introspective access not yet born, Donders took for granted the logic underlying such a method—that mental states could be inferred on the basis of objective patterns of responses rather than relying on asking the subject the question, “Which of these two tasks is easier?”

Time as the variable to estimate the nature of mental computation underlies dozens of methods: the Stroop task, episodic or repetition priming, semantic priming, evaluation priming, and many others to assess attention, perception, memory and categorization. We focus on one such measure, the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), which provides an estimate of the strength of association between concepts and attributes, much like the semantic priming measure does (see Wittenbrink, Chapter 2, this volume). In a grant proposal submitted by Banaji and Greenwald on January 13, 1994, the logic underlying the IAT was described as follows:

Experiment 3.8: Measurement of implicit attitude (B: Rapid classification method). The same materials as Expt. 3.7 are used, but without priming. Instead, two categories of words are assigned to each of two response keys. Subjects are asked to rapidly press (say) the right key whenever the stimulus word is *either* female-associated or pleasant in meaning, and the left key for words either male-associated or unpleasant in meaning. Through the course of a session, blocks of trials with the four combinations of category pairings and key assignments are intermixed. . . . The measure of implicit attitude . . . is the difference between latency with pleasant/male pairing versus pleasant/female pairing. To the extent that responding is faster with pleasant/female than with pleasant/male pairing, the latency-difference

measure indicates greater positivity of the implicit attitude associated with female.

In 1995, Greenwald and Banaji argued that although many existing effects were already available as evidence of implicit social cognition, an individual difference measure with the sensitivity to detect variability among a population would be needed. They concluded their review of implicit social cognition with the following comment:

In summary of existing efforts at indirect measurement of implicit social cognition: Research on latency decomposition, projective tests, and miscellaneous other procedures indicate that indirect measurement of individual differences in implicit social cognition is possible. At the same time, such measurement has not yet been achieved in the efficient form needed to make research investigation of individual differences in implicit social cognition a routine undertaking. When such measures do become available, there should follow the rapid development of a new industry of research on implicit cognitive aspects of personality and social behavior. (p. 20)

The IAT technique was developed to facilitate such progress, and in particular to generate new methods that would demonstrate large effect sizes and maintain reasonable reliability (Greenwald et al., 1998; for a discussion of the origins of the IAT, see Dasgupta, Greenwald, & Banaji, 2003). The technique was developed in the mid-1990s, following several years of parameter testing and refinement in labs at the University of Washington and Yale University. Distribution to any interested scientist preceded its first publication in 1998.

In the 9 years since its publication, the IAT method has received significant attention. At present, more than 200 papers report use of the method, hundreds of conference papers have been delivered, and more than 4.5 million tests have been taken at www.implicit.harvard.edu. In addition, a healthy skepticism about what the method is and does has produced commentaries on interpretation (see Nosek, Greenwald, & Banaji, in press, for a review). Specific reports of interest to readers include those that summarize results obtained using the IAT (Carney, Nosek, Greenwald, & Banaji, in press; Greenwald & Nosek, 2001; Nosek, Banaji, & Greenwald, 2002a; Nosek et al., in press), provide details on method and scoring development (Greenwald, Nosek, & Banaji, 2003; Nosek, Greenwald, & Banaji, 2005), and discuss the unique nature of reactions to the IAT (Banaji, 2001). In addition, stimulus materials and sample programs are available via web sites (see www.people.fas.harvard.edu/~banaji, www.briannosek.com, www.faculty.washington.edu/agg).

With these resources already available, this chapter focuses on an aspect of the work that is currently unavailable in a single location—a brief introduction to those who are new to the IAT and wish to become educated users of the technique and consumers of research that uses it. If successful, the chapter will provide a user-friendly guide to the IAT.

OVERVIEW OF THE IAT AND ITS INTERPRETATION

Like the semantic priming method used to understand semantic memory, the IAT measures the relative strength of association between pairs of concepts, labeled for pedagogical purposes as *category* and *attribute*. When completing an IAT, participants rapidly classify individual stimuli that represent category and attribute (in the form of words, symbols, or pictures) into one of four distinct categories with only two responses. The underlying assumption is that responses will be facilitated—and thus will be faster and more accurate—when categories that are closely associated share a response, as compared to when they do not. To intuitively understand how an IAT works, one might visit www.implicit.harvard.edu and try a test.

Structure of the Seven-Block IAT

Figure 3.1 presents a schematic overview of the structure of the first published IAT report, which assessed implicit attitudes toward flowers, relative to insects (Greenwald et al., 1998). It is described in detail here because it forms the basis of the rest of the chapter. At the heart of the measure is a pair of target concepts—in this example, flowers and insects—and a pair of attribute concepts—in this example, good and bad.

In Stage 1 of this sample task, participants rapidly classify words into the categories flower (by pressing the left computer key) and insect (by pressing the right computer key). They then repeat the same task for the categories good and bad (Stage 2). In Stage 3, the previous two tasks are combined and participants press the left computer key when any item in the category flower or good appears on the screen, and press the right computer key when any item in the category insect or bad appears on the screen (abbreviated as the *flower + good* or *insect + bad* pairing). Stage 4 repeats this procedure with an additional set of trials. In the next stage, the task in Stage 2 is reversed. Similarly, Stages 6 and 7 reverse the earlier combined pairings of Stages 3 and 4: flower + bad now share the left response key, and insect + good share the right response key. Because attitudes toward

Stage	Left key assignment	Right key assignment
1	FLOWER	INSECT
2	GOOD	BAD
3	FLOWER	INSECT
	GOOD	BAD
4	FLOWER	INSECT
	GOOD	BAD
5	BAD	GOOD
6	FLOWER	INSECT
	BAD	GOOD
7	FLOWER	INSECT
	BAD	GOOD

FIGURE 3.1. Schematic overview of the Implicit Association Test.

flowers are anticipated to be more positive than attitudes toward insects for most people, participants are expected to respond more rapidly, on average, when the category labels flower and good share one response, and insect and bad share the other response (Stages 3 and 4), as compared to the stage in which flower and bad share one response, and insect and good share the other response (Stages 6 and 7).

The difference in the latency to respond to particular pairings of concept and attribute (say, *insect + good* and *flower + bad*), compared to another set of pairings of concept and attribute (*insect + bad* and *flower + good*), provides an index of the relative strength of association between the first versus the second pairings. If the first set produces faster responses overall than the second (and does so even when the pairings' presentation order is reversed), we conclude that the relative strength of association between *flower + good* and *insect + bad* is greater than that between *flower + bad* and *insect + good* and therefore reflects a relative implicit preference for flowers over insects.

When young boys, compared with young girls, show a weaker preference for flowers over insects, we conclude that the group difference reflects a meaningful distinction in automatic preferences for these attitude objects (Banaji, in press; Baron & Banaji, 2006). When entomologists show a smaller effect on the same test, compared to a control group (Citrin & Greenwald, 1998)—that is, demonstrate a weaker relative preference for flowers over insects—we conclude that the IAT reveals individual differences in strength of association, using a known groups validation approach. That is, we begin with groups

a priori expected to show differences in attitude and test this expected difference on the IAT measure. Furthermore, if individual differences in strength of association represent meaningful differences in attitude, the test should predict other behaviors such that those with stronger *insect + good* scores are more likely to spend time with insects and to act in favorable ways toward insects, such as feeding and caring for them and being invested in their survival. Taken together, these findings—that groups differ in strength of associations in predicted ways, that one's personal experience alters the magnitude of those associations, and that those associations relate reliably to other individual judgments and behaviors—would provide evidence of the usefulness of the measure to understand implicit social cognition.

Categories can represent any grouping, such as *insects* and *flowers*, *Ohio State* and *Michigan*, *psychology* and *chemistry*, *elderly* and *young*, *Mac* and *PC*, or *Coke* and *Pepsi* (see Nosek, 2005, for a list of over 50 pairings that have been tested and interpreted). The attribute concept can also vary in many ways. If attitude or preference is of interest, the attribute dimensions can be represented by the labels *good/bad*, *pleasant/unpleasant*, or *positive/negative*.

Alternatively, the association between a target category (*female* or *male*) and a specific trait or attribute provides a measure of what is commonly thought of as a *belief* or *stereotype*. For example, the group *male*, compared to the group *female*, is more strongly associated with *mathematics*, relative to *liberal arts* (Nosek, Banaji, & Greenwald, 2002b), *career* relative to *family* (Nosek et al., 2002a), and *strong* relative to *weak* (Rudman, Greenwald, & McGhee, 2001), indicating implicit gender stereotypes.

Moreover, the attribute dimension can be turned toward the *self*; a measure of association between the categories (say, PC/Mac) and self/other can be obtained as a measure of relative *identity* with the objects. For example, the pairing of self with gender provides a measure of gender identity (see, e.g., Greenwald & Farnham, 2000), of self with ethnic group offers a measure of ethnic identity (Devos & Banaji, 2005), and of self with math/arts provides a measure of an academic identity (Nosek et al., 2002b).

Likewise, association of self as the target category and attitude as the attribute dimension is assumed to provide a measure of implicit self-esteem. It is the relative strength of association between *self* (compared to another category such as *other*) and positive/negative concepts. The IAT to measure self-esteem is structured like the attitudinal measure described above, with *self* and *other* serving as the target concepts and *good* and *bad* terms as the attribute dimension (e.g., Bosson, Swann, & Pennebaker, 2000; Dijksterhuis, 2004; Greenwald & Farnham, 2000; Karpinski, 2004; Yamaguchi et al., 2006).

The IAT's main usage has been in the domain of implicit social cognition, and hence most of our examples reflect such practice. However, the IAT had been used in several other contexts, such as clinical (e.g., Egloff & Schmukle, 2002; Gemar, Segal, Sagrati, & Kennedy, 2001), developmental (Baron & Banaji, 2006), marketing (Maison, Greenwald, & Bruin, 2001), and health (Czopp & Monteith, 2003) applications, as well as in legal scholarship (Kang & Banaji, 2006) and business (Banaji, Bazerman, & Chugh, 2003). For example, clinical psychologists have adapted and used the IAT to understand the mechanisms that differentiate groups of individuals with and without particular disorders, in illuminating the cognitive changes that occur during psychological treatment, and in predicting the likelihood of engaging in behaviors that are known to be associated with clinical disorders (Egloff & Schmukle, 2002; Gemar et al., 2001; Gray, Brown, & MacCulloch, 2005; Gray, MacCulloch, Smith, Morris, & Snowden, 2003; Jordan, Spencer, Zanna, Hoshino-Browne, & Correll, 2003; Nock & Banaji, 2006; Teachman, Gregg, & Woody, 2001; Teachman & Woody, 2003). Domains as diverse as identity with nature (Schultz, Shriver, Tabanico, & Khazian, 2004), attitudes toward death (Bassett & Dabbs, 2003), and toward celebrities, foods, cities and geography, public opinion issues, and politics, have been studied (Nosek, 2005).

In this chapter, we review over 100 published and in-press studies that include at least one IAT. In order to decide if an IAT method is necessary or worth using, the reader must first be aware of the main results associated with the measure and the current status of research on the reliability and validity of the measure. With this in mind, we divide the chapter into three sections that present (1) an overview of the task and a summary of the most robust findings associated with it; (2) an overview of the evidence for the reliability and validity of the task; and (3) a how-to guide for researchers wanting to use the IAT themselves or to better evaluate experiments using the IAT.

THE IAT REVEALS ROBUST ASSOCIATIONS THAT OFTEN DIFFER FROM SELF-REPORTED ATTITUDES AND BELIEFS: EVIDENCE FROM WEB-BASED SAMPLES

By measuring associations between concepts and attributes, the IAT can reveal associations that often differ from those that are introspectively accessed and reported verbally. Table 3.1 summarizes the results of 17 different IATs, based on over 2.5 million tests completed at two public web sites. To facilitate comparison between implicit

and explicit measures, means are reported as Cohen's *d* effect sizes of the difference from "no preference." At each site, visitors have an opportunity to complete one or more tests of their choice and receive feedback about the strength of their automatic associations.¹ Participants select a task, provide optional demographic data, and respond to explicit questions that parallel the IAT measure. Because of the large number of participants and the broad demographic range of respondents (compared to the typical undergraduate sample), these data provide insights into the variability among cognitions, offer estimates of correlations between implicit and explicit measures, and allow exploration of patterns of associations across different demographic groups. In addition, such large data sets can generate knowledge of the psychometric properties of the IAT and improve methodological and analytical techniques.

Two main findings are apparent from these data. First, subjects demonstrated strong and robust associations between social groups and basic evaluation (implicit attitude) or specific qualities (implicit stereotypes). Implicit attitudes toward culturally valued groups were shown to be positive; participants demonstrated, on average, greater positivity for White over Black, Other Peoples (non-Arab Muslims) over Arab Muslims, abled over disabled, young over old, and straight over gay. In addition, participants showed stereotype-consistent associations between White and American, male and science, and Black and weapons. These findings are consistent with laboratory data that used the IAT to measure the same constructs (e.g., Devos & Banaji, 2005; Greenwald et al., 1998, 2002; Hummert, Garstka, O'Brien, Greenwald, & Mellott, 2002; Jellison, McConnell, & Gabriel, 2004; Nosek et al., 2002b; Steffens & Buchner, 2003).

The second clear message from the web data is that patterns of cognitions can vary widely across implicit and explicit measures. For example, although participants showed strong implicit preference for White over Black (average Cohen's *d* of two race attitude tasks = 0.80), the effect of their self-reported bias was much weaker (average Cohen's *d* of two race attitude tasks = 0.31). It is not the case that the implicit measures always detect greater negativity than explicit measures, but in the case of attitudes toward social groups, that is a common result.

RELIABILITY AND VALIDITY OF THE IAT

Evaluating the IAT's validity is a somewhat different undertaking than that of evaluating the validity of a self-report scale. Because the

TABLE 3.1. Implicit and Explicit Attitudes for 17 Tasks Completed between July 2000 and May 2006 at Publicly Available Websites

Task	Higher numbers reflect association between:	Dates administered		IAT		Explicit		IAT-Explicit	
		Begin	End	N	D	N	d	N	d
Age attitude	Young + Good/Old + Bad	7/2000	5/2006	351,204	0.49	1.23	356,308	0.51	.13
Race attitude	White + Good/Black + Bad	7/2000	5/2006	732,881	0.37	0.86	759,566	0.36	.31
Skin-tone attitude	Light skin + Good/ Dark skin + Bad	3/2001	5/2006	122,988	0.30	0.73	72,735	0.25	.22
Child-race attitude	White + Good/Black + Bad	11/2001	5/2004	28,816	0.33	0.73	41,886	0.15	.29
Arab-Muslim attitude	Other Peoples + Good/Arab Muslims + Bad	11/2001	5/2006	77,254	0.14	0.33	37,499	0.58	.34
Religion attitude	Christian + Good/Jewish + Bad	8/2003	5/2006	66,092	-0.15	-0.34	43,711	-0.13	.38
Disability attitude	Abled + Good/Disabled + Bad	6/2003	5/2006	38,544	0.45	1.05	23,120	0.57	.14
Sexuality attitude	Straight People + Good/Gay People + Bad	2/2002	5/2006	269,683	0.35	0.74	168,498	0.54	.43
Weight attitude	Thin + Good/Obese + Bad	3/2001	5/2006	199,329	0.35	0.83	110,548	0.88	.20
Race-weapons stereotype	White + Harmless Objects/Black + Weapons	11/2001	5/2006	85,742	0.37	1.00	91,771	0.31	.15
American-native stereotype	White Am. + American/ Native Am. + Foreign	5/2002	5/2006	44,878	0.23	0.46	41,966	-0.42	.18

(Continued)

TABLE 3.1. (continued)

Task	Higher numbers reflect association between:	Dates administered		IAT		Explicit		IAT-Explicit	
		Begin	End	N	D	N	d	N	r
American-Asian stereotype	White + American/ Asian + Foreign	3/2001	5/2006	57,569	0.26	27,734	0.62	0.45	.17
Gender-science stereotype	Male + Science/ Female + Liberal Arts	7/2000	5/2006	299,298	0.37	139,182	0.93	0.79	.22
Gender-career stereotype	Male + Career/ Female + Family	10/2002	5/2006	83,084	0.39	82,550	1.10	0.89	.16
Presidential attitude	Bush + Good/ Other President + Bad ^a	5/2003	5/2006	68,123	-0.07	73,595	-0.15	-0.73	.54
Election 2004 attitude	Bush + Good/ Gore + Bad	10/2003	5/2005	22,904	-0.14	21,421	-0.27	-0.42	.71
Election 2000 attitude	Bush + Good/ Kerry + Bad	7/2000	2/2003	27,146	-0.09	29,925	-0.16	-0.20	.75
Total or median				2,575,535	0.33	2,122,015	0.73	0.36	.22

Note. From Nosek et al. (2006). D represents the IAT score calculated according to the recommendations of Greenwald et al. (2003).
^aThe comparison category to President Bush varied.

IAT represents a procedural format for measuring implicit cognition rather than a single measure of a specific construct, there is no single incantation of the IAT to be validated. Unlike traditional measures of attitude, such as the Modern Racism (McConahay, 1986) or Rosenberg Self-Esteem (Rosenberg, 1989) scales, and given that the IAT can be adapted to measure the constructs of stereotypes, self-esteem, identity, and attitudes toward concepts as diverse as gender, ethnicity, academic domains, and favorite foods, two IATs may have little in common other than the basic structure of the task. Thus, there are both general (format-specific) and specific (construct-specific) issues of reliability and validity to contend with in evaluating the IAT. By looking at issues of reliability and validity across content areas, we focus on the psychometric properties of the task in general. If IATs across multiple studies and multiple designs produce consistent effects, despite the variance associated with idiosyncratic design features, the generalizability of the findings should inspire greater confidence.

However, simply because the IAT format produces a reliable and valid task does not mean that any single IAT is necessarily a good measure of the target construct. Two IATs that measure attitudes toward the same construct may vary widely—one may use picture stimuli, and one may use verbal stimuli, or the exemplars of the category and attribute may differ quite a bit across task. As a result, features specific to construction of a particular IAT can produce unique variance.

Reliability

Error variance is easily introduced in response latency studies—a subject's sneeze, a car horn, or even an eyeblink, that coincides with the appearance of the stimulus can introduce task-irrelevant variability in response latency. Such factors are not typically given much importance in gauging the reliability of self-report scales, but they may dampen the reliability and stability of a measure based on quick responses. Indeed, the internal consistency of measures based on response latency is generally somewhat lower than that of those based on self-reports (Buchner & Wippich, 2000; Perruchet & Baveux, 1989).

In one study examining the internal consistency of a number of implicit measures, the IAT showed reasonable reliability (Cronbach's $\alpha = 0.78$); notably, this was higher than that of an evaluative priming task and a modified version of the IAT included in the same session (Cunningham, Preacher, & Banaji, 2001). Additional studies that have reported the internal consistency of the IAT also indicate

that it is generally acceptable (e.g., internal reliabilities averaged .79 across 50 studies in a meta-analysis conducted by Hofman, Gawronski, Gschwendtner, Le, & Schmitt, 2005).

The IAT has also shown reasonably good reliability over multiple assessments of the task. As shown in Table 3.2, in 20 studies that have included more than one administration of the IAT, test–retest reliability ranged from .25 to .69, with mean and median test–retest reliability of .50. Notably, in an analysis of seven implicit measures of self-esteem, the IAT’s test–retest reliability (.69) was superior to the other implicit measures of self-esteem, which ranged from –.05 (Stroop task) to .63 (initials–birthday preference task) and averaged .30 (Bosson et al., 2000).

Validity

The traditional “multitrait–multimethod” (MTMM) (Campbell & Fiske, 1959) approach dictates not only that IATs measuring similar constructs should correlate with one another, but also that IATs assessing theoretically distinct concepts should not. It is important that, across methods, measures of similar traits should also converge with one another but measures of distinct traits should diverge from one another. In addition, multiple IATs assessing distinct constructs relate to each other in theoretically predicted patterns, providing evidence for the IAT’s nomological validity. Evidence for its convergent validity with other implicit measures is more mixed. Discriminant validity is seen in studies indicating that cognitions that are predicted to be unrelated do, in fact, diverge from one another. That is, predicted relationships among different IATs are observed.

Nomological Validity: Theoretically Predicted Results Emerge across Studies

Greenwald and colleagues (2002) argued that the tendency toward cognitive consistency—as described almost a half century ago by balance (Heider, 1958) and dissonance (Festinger, 1957) theories—leads to cognitive balance among attitudes, stereotypes, identities, and self-esteem. For example, the higher a person’s self-esteem, the more positive ingroup attitudes he or she should have in order to maintain cognitive consistency. This can be phrased as, “If I am good, and I am strongly tied to my group, then my group is good.” Indeed, self-concept and attitudes related to gender and race (as measured by the IAT) adhered to this pattern, whereas explicit measures of these constructs did not. Similarly, the more women implicitly identified with

TABLE 3.2. Test–Retest Correlations of the IAT

Authors	Construct	Time period	<i>r</i>	<i>N</i>
Banse et al. (2001)	Attitudes toward homosexuality	Same session	.52	101
Bosson et al. (2000)	Self-esteem	4 weeks	.69	80
Cunningham et al. (2001) ^a	Racial attitudes	2 weeks between each of four sessions	.32	93
Dasgupta & Asgari (2004)	Gender stereotypes	1 year	.25	52
Dasgupta et al. (2001)	Racial attitudes	24 hours	.65	48
Dasgupta, McGhee, & Greenwald (2000)	Racial attitudes (name versus picture IAT)	Same session	.39	75
de Jong et al. (2003)	Fear association with spiders	4 months	.41	37
Egloff & Schmukle (2002)	Anxiety identity	1 week	.57	41
Egloff et al. (2005)	Anxiety identity	1 week	.58	65
		1 month	.62	39
		1 year	.47	36
Greenwald & Farnham (2000)	Self-esteem Two variants: an affective and an evaluative version	Same session	.43	145
		Same session	.68	58
	Self-esteem Two variants: a generic and an idiographic version	8 days	.52	44
Schmukle & Egloff ^b (2004)	Anxiety identity	Same session	.50	45
Shultz et al. (2004)	Implicit identity with nature	Same session	.45	32
		1 week	.46	33
		4 weeks	.40	33
Steffens & Buchner (2003)	Attitudes toward gay men	1 week	.50	84
		10 minutes	.52	107
Average			.50	
median			.50	

^aMean test–retest correlation of four IATs, each administered 2 weeks apart.

^bReflects control condition only. In the experimental condition, which was designed to elicit change in anxiety identity, participants completed a public speaking task in between administrations of the IAT. Correspondence between the two IAT scores was lower in this condition, $r = .21$.

their gender group and associated female with liberal arts (compared to math), the more they implicitly liked liberal arts (Nosek et al., 2002b). Students were also more likely to show implicit preference for their university over its main competitor to the extent that they showed both high implicit self-esteem and strong implicit university identity (Lane, Mitchell, & Banaji, 2005).

Cunningham, Nezlek, and Banaji (2004) examined whether implicit biases toward a range of stigmatized groups were related to one another. Such a pattern is predicted by traditional and modern theories of ethnocentrism and would therefore provide additional evidence of convergent validity. They found that attitudes toward five different stigmatized groups (involving race, sexuality, social class, religion, and nationality) loaded on a single factor of “implicit ethnocentrism.” It is important to note that a conceptually unrelated IAT (measuring attitude toward trees and birds) did not load on this factor.

Convergent Validity: The Relationship of the IAT to Other Implicit Measures

Patterns of convergent validity among implicit measures that purport to measure that same construct are more mixed. Several studies have shown little overlap between implicit measures designed to assess the same construct. Bosson and colleagues (2000) examined the correlations among seven implicit and four explicit measures of self-esteem. Although the IAT was uncorrelated with the other implicit measures, it was not alone in failing to converge—among the 15 possible zero-order correlations between the six measures, only two pairs of implicit measures significantly correlated with each other. Most attention has focused on understanding when and how the two most widely used measures of implicit attitudes, evaluative priming tasks and the IAT, converge and diverge. In a typical use of the priming task measuring racial attitudes (Fazio, Jackson, Dunton, & Williams, 1995), participants classified words as positive or negative as quickly as possible. An image of a Black or a White face preceded each word, allowing the assessment of the relative facilitation of each social group to positive or negative concepts. The viewing of Black faces, as compared to White faces, facilitated judgments of negative words and interfered with judgments of positive words, indicating that Black faces automatically activated negative concepts.

IATs and priming tasks measuring implicit attitudes toward smoking (Sherman, Rose, Koch, Presson, & Chassin, 2003) and condom use (Marsh, Johnson, & Scott-Sheldon, 2001) were unrelated.

Fazio and Olson (2003) reported that across four studies in their lab, priming tasks and IATs measuring racial attitudes did not correlate. Other studies show greater promise. Stereotypes about gender authority (associations between female and low-status jobs), as measured by the IAT, were correlated with three indices of attitudes toward female authorities derived from the priming task (Rudman & Kilianski, 2000). Moreover, participants who tended to show strong implicit gender stereotypes (on the IAT) also showed more positive attitudes toward women on the priming measure, and political attitudes measured by a priming task and the IAT were reliably related (Nosek & Hansen, 2004).

What accounts for these sometimes less-than-robust correlations? As mentioned earlier, response-latency measures often have attenuated internal reliability, and on occasion this has been especially true of priming (Bosson et al., 2000). As a result, true relationships between measures may be masked by measurement error. Three implicit measures of racial preference (a standard IAT as well as response-window versions of both the priming task and IAT in which participants made responses in a very short window and that used error rates as the dependent measure) were completed during two sessions separated by a 2-week interval (Cunningham et al., 2001). Latent variable analysis revealed that this approach improved the stability of the measures. Moreover, the measures were correlated: Two versions of the IAT were strongly related, $r = .77$, as was the priming task with both the response-window IAT, $r = .53$, and the standard IAT, $r = .55$. In addition, all three implicit measures loaded onto a single "implicit bias" factor that was distinct from, but strongly correlated with, explicit bias. These findings suggest that when reliability is accounted for, implicit measures are more likely to be related.

Schwarz (1999) pointed out that many features of an explicit scale, such as the order of questions, response options, or slight wording changes, can affect the provided responses. Similarly, different implicit measures may tap into different features of an attitude object. One of the major distinctions between the priming task and the IAT is that the IAT requires explicit categorization by race, whereas the priming task does not. Noting this, Olson and Fazio (2003) suggested that participants completing the priming task evaluate exemplars of a group, whereas participants completing an IAT evaluate the overall social category. To support this contention, an IAT and a priming task measuring racial attitudes covaried only when participants completing the priming task were instructed to categorize the prime faces by their race; that is, when the subjects' task was made more similar to the IAT. However, this change also in-

creased the split-half reliability of the priming task from .04 to .39 (Olson & Fazio, 2003), leaving open the possibility that the null relationship between the IAT and standard priming task was due to the relatively lower reliability of the standard priming task.

Given that correlations can be drastically attenuated when measures are unreliable, it is difficult to interpret null relations when one of the measures shows poor reliability. Based on the available evidence, we expect that additional investigations that use large samples, maximize reliability, and correct for measurement error through latent variable analysis will clarify the nature of the relationship between implicit measures.

Discriminant Validity

Overlap between IATs that are conceptually related would be poor evidence of the IAT's validity if measures that are conceptually distant also positively correlated with one another. For example, in the Cunningham and colleagues (2004) study, only attitudes toward social objects were related, consistent with theories of ethnocentric bias. The nonsocial attitude did not load onto the factor of "implicit prejudice" in their analysis. Their findings support the idea that implicit attitudes that are expected to be related do converge, whereas those that are conceptually distinct diverge.

Other studies, however, have found correlations between conceptually distinct IATs (McFarland & Crouch, 2002; Mierke & Klauer, 2003). Undoubtedly, as with any measure, there is variance in an IAT score that is attributable to an individual tendency to show IAT effects. Mierke and Klauer (2003) found that an IAT assessing novel associations covaried with both a flower–insect IAT and a shyness–self-concept IAT. Notably, after subjecting their data to the most recent suggested scoring procedures (Greenwald et al., 2003) for the IAT, method-specific variance was either removed (Study 2) or "markedly reduced" (p. 1188; see Back, Schmukle, Egloff, & Gutenberg, 2005, for a similar result). These findings suggest that when using statistical methods that better account for method variance—such as latent variable analysis or improved IAT scoring procedures, better discriminant validity is likely to be observed.

Using a method more directly analogous to the MTMM approach, Gawronski (2002) measured implicit and explicit attitudes toward Turks and Asians among German participants using the IAT. If the two IATs were tapping an individual difference in group-specific attitudes, he reasoned that the IAT measuring attitudes toward a particular group should relate only to explicit attitudes to-

ward that group. This is the pattern that emerged—the IAT measuring attitudes toward Asians correlated only with explicit attitudes toward Asians, whereas the IAT measuring attitudes toward Turks correlated only with explicit attitudes toward Turks. Similarly, across seven attitude objects, Nosek and Smyth (in press) showed that, generally, each IAT-based measure of attitude correlated with an explicit measure of attitudes toward that object, but not with explicit attitudes toward the other target objects. Furthermore, while the IAT and explicit measures were related, they also retained unique components that were not reducible to shared method variance.

Criterion Validity

This section briefly summarizes research indicating the IAT's performance on tests of criterion validity. In short, the IAT can predict group membership based on theoretically predicted patterns of ingroup attitudes and identification, correlates with (but is distinct from) explicit measures of associated constructs, and successfully predicts judgments and behaviors.

Known-Groups Validity. If a new test is designed to be a valid test of math knowledge, then math majors should outperform non-math majors on it. This “known-groups” approach to validity argues that a good measure should reliably distinguish between members of different groups, based on a priori predictions or knowledge about those groups. The IAT has indeed demonstrated theoretically predicted patterns of strong ingroup liking. Japanese Americans exhibited strong preference for their group relative to Korean Americans, whereas Korean Americans showed the opposite pattern (Greenwald et al., 1998). Similarly, East and West Germans each exhibited preference for their ingroup (Kuhnen et al., 2001), and even members of groups artificially created in the laboratory showed preference for their ingroups (Ashburn-Nardo, Voils, & Monteith, 2001). Men and women associated their own gender strongly with self (Aidman & Carroll, 2003; Greenwald & Farnham, 2000), and women, consistent with the prevailing social stereotype, implicitly preferred the arts to math more than men did (Nosek et al., 2002b).

System justification theory (SJT) (Jost & Banaji, 1994) makes a more subtle theoretical prediction about ingroup preference. Specifically, SJT predicts that members of lower-status groups should show reduced implicit liking for their ingroup (compared to members of higher-status groups). Despite the ubiquity of ingroup preference, but consistent with SJT, the IAT is sensitive to differences in the soci-

etal evaluation of different groups (see Jost, Banaji, & Nosek, 2005, for a review). On the IAT, Black Americans showed reduced ingroup preference as compared to Whites (Ashburn-Nardo, Knowles, & Monteith, 2003; Livingston, 2002; Nosek et al., 2002a), and overweight and poor people (Rudman, Feinberg, & Fairchild, 2002) actually showed outgroup preference. Status also moderated the strength of ingroup preference among students at universities that varied in prestige (Jost, Pelham, & Carvallo, 2002), and lower ingroup preference has been shown in young children from disadvantaged groups, suggesting the early learning of justifying tendencies that are visible on implicit but not on explicit measures (Dunham, Baron, & Banaji, 2006b).

Successful discrimination between group members even extends to “groups” that are defined by behavior rather than demographic traits. In a study of subjects who were snake or spider phobic (Teachman & Woody, 2003), a composite measure of three IATs successfully classified 92% of participants according to which creature they feared. Smokers showed more positive attitudes toward and stronger identity with smoking than nonsmokers (Swanson, Rudman, & Greenwald, 2001). Although both light and heavy drinkers held negative implicit attitudes toward alcohol (as compared with soda), heavy drinkers strongly associated alcohol and arousal, whereas light drinkers did not (Wiers, van Woerden, Smulders, & de Jong, 2002). Gray and her colleagues (2003) recently investigated implicit attitudes about violence among psychopathic and nonpsychopathic murderers. Although all groups showed a preference for the concept peaceful, as compared to the concept violent, psychopathic murderers showed less dislike for violence than did nonpsychopathic murderers. In addition to providing evidence that the IAT can discriminate even between different types of deviance, this finding was taken to suggest that violent acts committed by psychopaths may be rooted in unusual beliefs about violence, unlike violent acts committed by nonpsychopaths, which may stem from other causes. In each of these cases, it is unclear whether the automatic cognitions influenced the subsequent behaviors, whether cognitions changed because of behavior, or both.

Relationship with Explicit Measures. The nature of the relationship between implicit and explicit attitudes has received a great deal of attention that has not answered the original proposed question: “Do implicit and explicit attitudes relate to one another?” As noted by Fazio and Olson (2003), the more appropriate question may be, “Under what conditions, and for what kind of people, are implicit

and explicit measures related?” (p. 304). Useful answers may emerge from questions focused on the conditions under which implicit and explicit measures will covary.

There is a wide range in the extent to which implicit and explicit attitudes covary. As seen in Table 3.1, across 17 IATs that were available at public websites, correlations between implicit and explicit measures ranged from $r = .13$ to $r = .75$ (median $r = .22$). Laboratory studies have shown similar variability, with a number of studies revealing only slight or moderate (but generally positive) correlations between the IAT and explicit measures of the same construct (e.g., Bosson et al., 2000; de Jong, van den Hout, Rietbroek, & Huijding, 2003; Egloff & Schmukle, 2002; Greenwald et al., 1998; Karpinski & Hilton, 2001; Ottaway, Hayden, & Oakes, 2001; Rudman & Kilianski, 2000) and other studies showing strong and robust correlations between the IAT and explicit measures (e.g., Asendorpf, Banse, & Muecke, 2002; Cunningham et al., 2001; Greenwald & Farnham, 2000; Jellison et al., 2004; McConnell & Leibold, 2001; Wiers et al., 2002). A recent meta-analysis of such studies found that across 126 independent correlations, implicit–explicit correspondence ranged from $r = -.25$ to $r = .60$, with an average implicit–explicit correlation of .19 (Hofmann et al., 2005).

Consistent with the notion that implicit and explicit attitudes are distinct constructs (Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooler, 2000), even when the IAT and explicit measures do correlate, implicit and explicit attitudes are separate constructs. Confirmatory factor analyses indicated that self-esteem and gender identity were better fit by a model in which implicit and explicit measures loaded onto two separate constructs, rather than a single, latent construct (Greenwald & Farnham, 2000). This finding parallels that of Cunningham and colleagues (2001), in which implicit (IAT and priming) racial attitudes were distinct from, but positively correlated with, explicit racial attitudes. This conclusion generalized to a wide variety of domains—across 57 different pairs of attitude objects, a two-factor solution fit much better than a single-factor solution even when implicit and explicit attitudes were highly correlated with one another (Nosek, 2005; Nosek & Smyth, in press). Further support for the distinction between implicit and explicit attitudes comes from findings that implicit and explicit attitudes predict unique variance in meaningful criterion variables (see, e.g., McConnell & Leibold, 2001; Nosek et al., 2002b).

Given that the extent to which implicit and explicit attitudes are correlated varies widely across studies, more recent work has turned to the issue of the conditions under which implicit and explicit atti-

tudes will covary. Nosek (2005; see also Nosek & Banaji, 2002) identified factors that moderate the nature of the relationship between implicit and explicit attitudes: (1) self-presentational concerns, (2) attitude strength, (3) attitude dimensionality (the extent to which liking for one category appears to imply disliking of the contrasting category), and (4) attitude distinctiveness (the perception that an individual's attitude is distinct from others' attitudes). Also, consistent with the finding that implicit–explicit correlations will be higher for strongly held attitudes, participants who elaborated about an attitude or reported that an attitude was extremely important to them showed greater implicit–explicit correspondence than those who did not (Karpinski, Steinman, & Hilton, 2005).

This approach of examining aspects of the attitude object, as well as the attitude holder, when investigating the relationship between implicit and explicit measures will likely reveal additional personal and situational moderators of the relationship between implicit and explicit measures. The hunt for correlations may be most successful when large samples and rigorous statistical techniques, such as meta-analysis or latent variable modeling, are used, as these techniques are likely to provide the most reliable and stable correlations (be they negative, positive, or null).

The IAT Predicts Meaningful Behavior. Given psychologists' long-standing interest in understanding how attitudes predict behavior (Kraus, 1995; LaPierre, 1934), the pursuit of the IAT's ability to predict "real" behavior should not be surprising. In addition to successfully discriminating between groups of people who perform a behavior and those who do not (such as smoking, or avoidance of spiders), the IAT successfully predicts behavior. Poehlman, Uhlmann, Greenwald, and Banaji (2005) meta-analyzed 86 independent samples and found that the IAT predicted a range of criterion variables, including social judgments, physiological responses, and social action. In this section, we review some of the evidence that the IAT predicts behaviors and judgments in the domains that have received the most attention: stereotyping and prejudice, and health-related behaviors, such as food choices, alcohol use, and smoking.

Just as the first wave of research using the IAT centered on stereotyping and prejudice, the greatest focus in the attitude–behavior arena has been on behavior in intergroup settings. Implicit bias measured by the IAT predicts individual differences in behaviors and judgments. Stronger implicit stereotyping of Blacks covaried with more negative judgments of ambiguous actions by a Black target (Rudman & Lee, 2002; see Gawronski, Geschke, & Banse,

2003, for a similar result with a Turkish target). More negative attitudes toward Blacks (as compared to Whites) successfully predicted more negative nonverbal behaviors (e.g., less speaking time, less smiling, more speech errors) during an interaction with a Black experimenter (as compared to an interaction with a White experimenter; McConnell & Leibold, 2001). Similarly, spontaneous avoidance tendencies toward people with AIDS covaried with stronger negativity toward people with AIDS (as compared to healthy people; Neumann, Hulsbeck, & Seibt, 2004). Most recently, Green, Carney, Pallin, Iezzoni, and Banaji (2006) found that doctors with stronger anti-Black attitudes and stereotypes were less likely to prescribe thrombolysis for myocardial infarction to African American patients diagnosed with the same condition as equivalent White Americans.

These studies focused on the prediction of behavior toward an outgroup; the IAT also effectively predicts behavior toward the ingroup. Greater anti-Black sentiment predicted Blacks' preference for a White partner over a Black partner on an anticipated intellectually challenging task (Ashburn-Nardo et al., 2003). Among gay men, more positive attitudes toward homosexuality on the IAT were related to more positive experiences in the gay community (Jellison et al., 2004). Stronger implicit romantic fantasies (the implicit association between romantic partners and chivalry and heroism) were linked to women's reported interest in pursuing powerful activities, such as attaining education or high-status jobs (Rudman & Heppen, 2003).

In addition to these macro-level behaviors, the IAT also predicts lower-level perceptual and cognitive events. The utility of the IAT to predict unobtrusive perceptual tasks and uncontrollable physiological measures suggests that more negative implicit attitudes toward a group leads to more top-down stereotypic processing. In a series of striking demonstrations, Hugenberg and Bodenhausen (2003) found that negativity toward Blacks on the IAT predicted a lowered threshold for detecting hostility on Black, but not White, faces. The reverse effect also held—subjects had a lowered threshold for judging racially ambiguous faces with hostile expressions as Black (Hugenberg & Bodenhausen, 2004). In addition, more negative attitudes may result in the depletion of cognitive resources when facing a member of the target group: After an interaction with a Black confederate, White participants with stronger anti-Black bias showed more cognitive decrements than participants with lower anti-Black bias, as measured by performance on a Stroop task (Richeson & Shelton, 2003). Moreover, the extent of activation in the right dorsolateral prefrontal

cortex (DLPFC)—a brain region believed to be critical for executing cognitive control—when presented with unfamiliar Black faces was correlated with IAT scores and mediated the amount of interference on the Stroop task following interaction with a Black individual (Richeson et al., 2003). Other research using physiological measures points to the relationship between implicit intergroup attitudes and emotion. The IAT successfully predicted greater activation of the amygdala—an area of the brain associated with emotional, particularly fear, responses—to the presentation of unfamiliar Black (versus White) faces (Cunningham, Johnson, Gatenby, Gore, & Banaji, 2003; Phelps et al., 2000). Subsequent research (Cunningham et al., 2004) found that the IAT–amygdala relationship is stronger for subliminal than for supraliminal presentation of faces, indicating that the IAT reflects more automatic rather than controlled reactions to social groups.

Explicit reports of attitudes have been notoriously uninformative in accurately gauging health behaviors. Statistics regarding obesity, smoking, and sexually transmitted disease suggest that behaviors are not always congruent with people’s best intentions to eat well, stop smoking, or practice safe sex. Although there are undoubtedly many reasons for this discrepancy, one possibility is that implicit processes play a role in determining behavior. These behaviors may be especially likely to be influenced by implicit mechanisms, because they may be susceptible to self-presentational concerns and often happen in the “heat of the moment” (such as the decision to use or not use a condom during sex) or in situations prone to low inhibition (such as at a bar with freely flowing alcohol).

Despite a widely reported initial failure to find a relationship between IAT scores and a choice between a healthy (apple) and less healthy (candy bar) snack (Karpinski & Hilton, 2001), when the studies ensure sufficient power to detect an effect, a relationship between implicit associations and food choices is obtained: attitudes toward the more global categories *snacks* and *fruits* successfully predicted the choice of fruit or a less healthy snack at the end of the experimental session (Perugini, 2005). IAT measures of attitudes toward soda, relative to fruit juices, and high-calorie foods, relative to low-calorie foods, were also associated with self-reports of food consumption (Maison et al., 2001).

Use of the IAT by researchers interested in less adaptive consumptions has led to a better understanding of the cognitive processes guiding such choices. For example, research examining implicit associations related to alcohol use suggests that it may be the perceived positive effects of alcohol that distinguish heavy from light

drinkers. The IAT predicted the extent to which heavy drinkers showed arousal and the urge to drink in the presence of a glass of beer (Palfai & Ostafin, 2003). Positive, but not negative, associations with alcohol were significantly related to self-reports of alcohol use for the 30 days prior to the experimental session (Jajodia & Earleywine, 2003). Similarly, among heavy drinkers, stronger associations between the concept *alcohol* and the attribute *approach* (relative to *electricity* and *avoid*) were correlated with frequency of binge drinking and quantity of alcohol consumed per drinking session during the month leading up to the experimental session (Palfai & Ostafin, 2003). In addition to retrospective reports, the IAT also predicted alcohol use in the month following administration of the task (Wiers et al., 2002).

Similarly, the IAT has helped to better elucidate the factors leading people to smoke. Although smokers and nonsmokers equally disliked smoking when it was contrasted with a positive health behavior or an even more strongly stigmatized behavior (stealing), smokers showed greater positivity when the contrasting category was *not smoking* and were more strongly identified with *smoking* at the implicit level. Moreover, smoking identity related to number of cigarettes smoked per day (Perugini, 2005; Swanson et al., 2001). Positive implicit attitudes toward smoking among mothers predicted the likelihood that their children would smoke (Chassin, Presson, Rose, Sherman, & Prost, 2002).

The IAT Measure Shifts in Response to Situational Cues. In light of the preceding evidence, one might conclude that implicit attitudes are stable characteristics that do not vary. Such an approach would be in line with views of the implicit system as slow to learn associations, and consequently slow to change (Smith & DeCoster, 2000). However, a growing number of studies indicate that implicit attitudes, including those measured by the IAT, often shift in relation to the current situation and new learning (Blair, Ma, & Lenton, 2001; Dasgupta & Greenwald, 2001; Gregg, Seibt, & Banaji, 2006; Lowery, Hardin, & Sinclair, 2001; Richeson & Ambady, 2003; see Blair, 2002, for a review). For example, race bias (as measured by the IAT) decreased after participants viewed pictures of admired African Americans and disliked White Americans, and such effects persisted after a 24-hour period (Dasgupta & Greenwald, 2001). Using a new procedure of individuals modulating their own implicit attitudes, Akalis, Nannapaneni, and Banaji (2006) showed that self-generated thoughts in ordinary people and yoga practitioners shifted attitudes in both negative and positive directions. Implicit self-concepts also

appear susceptible to change: In a college-age sample, participants showed greater implicit identity with aggressive concepts after playing a violent video game than after playing a nonviolent video game (Uhlmann & Swanson, 2004).

Reconciling the findings that the IAT is able to predict meaningful criterion variables with findings that it is malleable may be an important avenue for future research. This area of inquiry will be especially important for applied researchers who desire a stable measure of implicit attitudes. The tendency for IAT-based attitudes or identities to shift in response to situational cues need not represent a challenge to its validity. Just as explicit attitudes may shift in response to the current situation, a particular situation may activate a specific set of associations, temporarily making certain category-target associations stronger. These associations, in turn, are reflected on the IAT. Understanding how the shifts in implicit cognitions relate to behavior may be an especially useful step in integrating these two lines of evidence.

FREQUENTLY ASKED QUESTIONS ABOUT THE IAT

This section briefly addresses some of the questions that are commonly asked of researchers using the IAT. Each of these questions has generated numerous studies and debates, and to fully address all of the issues raised by them is beyond the scope of this chapter. Here we outline the main issues involved in each matter.

- *Do IAT scores reflect attitudes of the individual or the culture?* In order for the IAT to function well in the purpose for which it was originally designed—measuring an individual difference in cognition—it needs to serve as more than a mirror of the culture. That is, it cannot be simply a tally of “the associations a person has been exposed to in his or her environment” (Karpinski & Hilton, 2001, p. 774), but should be useful in sorting among individuals within that culture. For instance, when a person shows a strong implicit preference for the Yankees over the Red Sox on the IAT, that person’s score should not primarily reflect “extrapersonal associations” (Olson & Fazio, 2004) that come from living in New York; it should indicate a propensity to root for the Yankees and should be linked to the association between the person and the Yankees. We agree. In fact, IAT-based attitude measures do reliably relate to an individual’s implicit cognitions about him- or herself (Greenwald et al., 2002) and also predict behavior (Poehlman et al., 2005).

If IAT scores were tapping environmental associations more than individual attitudes, they should correlate more reliably with self-reports of beliefs about widespread cultural preferences than with self-reports of a person's own preference. This is not the case; across 58 different attitude objects self-reported attitudes were consistently and reliably related to IAT performance, and estimates of the cultural attitudes such as beliefs about the "average person's" feelings were more related to self-reported attitudes than they were to the IAT (Nosek & Hansen, 2004). In fact, self-reported attitudes completely mediated the relationship between the IAT and perceptions of cultural attitudes.

- *Are implicit cognitions distinct from explicit ones? If so, are they the "true" attitudes, identities, or beliefs?* Just as implicit and explicit memory systems can diverge (Gabrieli, Fleischman, Keane, Reminger, & Morrell, 1995; Roediger, 1990, 2003), so too, it has been argued, can implicit and explicit attitudinal systems (Wilson et al., 2000). Although the IAT, like any other measure in this family, is not a process-pure measure—it seems to be affected by both automatic and controlled processes (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005)—data indicate that the construct measured by the IAT differs from that assessed by self-reported measures. Correlations between implicit and explicit attitudes vary from close to zero to .90. We take these relationships seriously, especially when subject to appropriate statistical constraints, in inferring the degree of overlap between implicit and explicit measures. The question of whether these measures tap into different underlying representations is not one that is easy to answer. We focus, rather, on the empirical result showing varying levels of overlap in the hope that these patterns of results over time will give an indication about differences in underlying representation.

The fact that participants are often surprised by their IAT scores (Monteith, Ashburn-Nardo, Voils, & Czopp, 2002; Monteith, Voils, & Ashburn-Nardo, 2001) suggests that the IAT taps attitudes or beliefs that are not accessible by conscious introspection. In addition, the IAT and explicit measures explain unique variance in math performance (Nosek et al., 2002b) and differ in their ability to predict behavior across domains (Poehlman et al., 2005). Furthermore, the discrepancy between the two sets of cognitions is meaningful—individuals with high explicit but low implicit self-esteem showed greater narcissism, exhibited more ingroup bias in a minimal group paradigm, and reduced dissonant attitudes more than those with high explicit and implicit self-esteem (Jordan et al., 2003). Finally, the finding that the IAT relates to amygdala activation more strongly

for faces presented subliminally than supraliminally suggests that it measures more automatic rather than controlled attitudes (Cunningham et al., 2004).

If these two cognitions—implicit and explicit—can exist simultaneously, is one of them the “real” one? To our knowledge, the only printed statements that the IAT measures a person’s “true” sentiment are those that argue against such a position (Arkes & Tetlock, 2004; Karpinski & Hilton, 2001). A person’s IAT score is no more a measure of his or her “true” attitude than that person’s response to a Likert scale. The elusive “true” attitude seems not to exist—when an attitude depends on the measurement context, mood of the subject, and prior questions, how does one decide which is the true one? Indeed, routes to explicit attitude change are so numerous that the topic merits its own *Handbook* (Petty, Wheeler, & Tormala, 2003) and *Annual Review of Psychology* (Petty, Wegener, & Fabrigar, 1997) chapters, and the nascent research on implicit attitudes shows similar sensitivity to contextual cues (Blair, 2002).

If one were going to use predictive ability to determine the “true” attitude, a similar dead end would emerge. Implicit and explicit attitudes better predict discriminatory and consumer behaviors, respectively (Poehlman et al., 2005), suggesting that the ability of each type of attitude to predict behavior depends largely on the topic being studied. Based on this evidence, it seems sensible to say that implicit and explicit attitudes are equally authentic possessions of their holders.

- *What processes underlie IAT effects?* Several lines of inquiry have investigated the psychological processes involved in the IAT, with researchers proposing mechanisms such as a random walk (Brendl, Markman, & Messner, 2001), figure–ground asymmetries (Rothermund & Wentura, 2001, 2004; see also Greenwald, Nosek, Banaji, & Klauer, 2005), stimulus–response compatibility (De Houwer, 2001), and task set switching (Klauer & Mierke, 2005; Mierke & Klauer, 2001, 2003). These efforts have provided insight, for example, into the stronger influence of category-level (relative to stimulus-level) representations in producing IAT effects (De Houwer, 2001; Olson & Fazio, 2003). In addition, many of these theory-driven approaches have spurred methodological changes, such as improved scoring procedures (Greenwald et al., 2003), or better understandings of effects of block order (Klauer & Mierke, 2005). Future research into the mechanisms that cause people to exhibit IAT effects will likely improve an understanding of its relationship with its cousin implicit measures and spur further methodological improvements.

HOW TO BUILD AN IAT

Selecting Appropriate Categories and Exemplars

On one hand, the IAT is flexible—its structure allows the researcher to measure a broad range of constructs with one tool. This same flexibility, on the other hand, could potentially lead to a temptation to pick any four categories and throw them into an IAT. This approach would ignore the fact that the IAT's structure constrains what constructs it can best capture. In this section, we review the different stages of IAT construction.

Categories Matter

Two obvious choices in developing an IAT arise in determining how to represent the chosen categories. Both chosen category labels, and the specific stimuli presented, determine this construal of the concept. Data suggest that it is the construal of the category that determines how it is evaluated.

All judgments are made in some context. The IAT forces the researcher to be explicit about defining this context, as the structure of the task can change the perception of the target object being evaluated. De Houwer (2001) measured attitudes toward *British* (relative to *foreign*) among British subjects. Exemplars of each category included three positive (e.g., Princess Diana, Albert Einstein) and three negative (e.g., Margaret Thatcher, Adolf Hitler) stimuli. The superordinate category, rather than the stimuli's valence, determined response latencies to each item. De Houwer concluded that category membership, rather than valence of individual exemplars, is most important in determining IAT effects. Similarly, Mitchell, Nosek, and Banaji (2003) varied the categorization tasks as to whether they were based on race or occupation. They found that participants preferred a set of (well-liked) Black athletes to (disliked) White politicians when categorized on the basis of occupation, but preferred the politicians to the athletes when categorization was based on race. Taken together, these findings imply that the nature and construal of the categories play a large role in determining IAT effects.

These results suggest that the first step in designing an IAT is to precisely define the constructs of interest, and this will influence the choices of category labels to represent the constructs. Many categories (e.g., male) have an obvious comparison category (e.g., female), and such category pairs lend themselves particularly well to use of the IAT and interpretation of its effects. In the case where there is no obvious comparison category, it may be desirable to use

an alternate implicit measure designed to measure single associations such as the Go/No-Go Association Task (Nosek & Banaji, 2001).

If the IAT is still the preferred method, a comparison category should be a sensible, mutually exclusive category that is ideally from the same domain (e.g., choosing *humanities* as the companion target category for the academic domain *science*), or presenting a category that represents the domain absent the target category (e.g., Arab Muslims, as compared to Other Peoples; see Table 3.1). Another alternative is to select an unrelated, neutral category. Some success has been realized in comparing social categories to (presumably) neutral categories such as “electricity” (Palfai & Ostafin, 2003) or “middle” (Lane & Banaji, 2004; Pinter & Greenwald, 2005), but such instantiations are not sufficiently understood to implement with confidence. One challenge is that no category is likely to be truly neutral, and interindividual variation in evaluation of the comparison category can introduce unwanted variance in effects. Second, the measurement effects of a relative comparison of a target category with an unrelated comparison are not understood.

Stimuli Matter

The popular press has suggested that citizens of some European nations (not to mention some Americans) love the nation America, but dislike its current administration (Bumiller, 2004). This idea would predict that America, when represented in an IAT by Dick Cheney and George Bush, would be evaluated negatively, but would be evaluated positively when denoted by pictures of the American flag and Golden Gate Bridge. In fact, stimulus exemplars do influence IAT effects. In addition to varying the category labels for Black athletes and White politicians, Mitchell and colleagues (2003) conducted a study in which they held the category labels constant and varied the stimuli used to represent the categories. When disliked Blacks and liked Whites represented the categories *Black* and *White*, participants exhibited strong and significant preference for Whites over Blacks. However, when liked Blacks and disliked Whites represented the categories, the typically seen preference for Whites was significantly diminished, with participants demonstrating nonsignificant preference for Whites (see also Govan & Williams, 2004; Steffens & Plewe, 2001). Similarly, inclusion of two images of lesbians as stimuli for the category *gay people* resulted in weaker implicit preference for straight over gay than when two images of gay men were included in the stimuli set (Nosek et al., 2005).

The prior examples represent changes in exemplars that were primarily designed to elicit different attitudes—by switching from liked Blacks/disliked Whites to disliked Blacks/liked Whites, the construal of the groups was changed and different patterns of preference were shown. Less drastic changes in category exemplars do not appear to shift implicit associations; when stimuli are chosen that do not alter the construal of the group, there are little to no effects of the chosen exemplars (Nosek et al., 2005). That is, if different stimuli represent the category in the same way, small differences among them are not likely to produce large differences in implicit attitudes, identities, or stereotypes. Stimuli that best represent the construal of the construct that a researcher is interested in will likely produce the most valid measure of implicit cognition.

For example, if a researcher assessing attitudes toward the category *Asian* were interested in attitudes toward Asian people, he or she would use faces easily identifiable as Asian, or names that can be quickly classified as Asian forenames or surnames. However, if the central research question focused on Asian culture or Asian nations, then a wider array of stimuli, including names of prominent cities or landmarks, would be appropriate. Note that the selection of stimuli does not guarantee that the construct is measured as intended.

Because the IAT relies on responses that are made without extensive deliberation, stimuli that are categorized easily and quickly will add the least error variance to the task. Pilot testing can ensure that participants can readily identify each item as denoting the appropriate category. Ambiguity about an item's appropriate categorization may slow reaction times, as may use of negations of words or phrases such as "unintelligent" that require additional time to be successfully negated and categorized correctly. Particularly when the IAT is used as an individual difference measure, these steps can help reduce task-related variability and maximize the variance the investigator cares about: that due to individual differences in the cognition.

Exemplars should be categorized solely on the basis of their membership in the appropriate category. That is, items should not be confounded with any of the other categories (Steffens & Plewe, 2001). An inadvertent confound—for example, all of the *good* words start with the letter C or are of more than seven letters, whereas all of the *bad* words start with the letter H or are of fewer than four letters—could provide subjects with a cue for sorting that is irrelevant to the task.

The addition of multiple cues for classification can reduce confusion or ambiguity about the task; for example, a researcher may choose to represent concept categories (such as *American* or *foreign*) with images, and attribute categories (such as *good* or *bad*) with

words. Category and attribute items can be distinguished even when all stimuli are text by using a particular font and color for the concept items and a different font and color for the attribute items.

Once the criteria for stimuli are determined, how many should there be? Greenwald and colleagues (1998) indicated that stimulus sets of 25 items and of 5 items produced implicit preferences of equal magnitude. In a more extensive test of the effects of the number of exemplars, Nosek and colleagues (2005) varied the number of exemplars in both the target and attribute categories on three different measures of implicit cognition (Black/White attitude, old/young attitude, and gender/science academic stereotype). Even with fewer than four stimulus items for each category, the overall magnitude of implicit biases was consistent, and smaller numbers of stimuli did not impair the reliability of the task, nor did it increase the influences of potential confounding variables. Only when the categories were represented by one or two items were the psychometric properties (correlation with self-reported attitudes and split-half reliability) reduced. Thus, it seems that better construct validity will be obtained when researchers select the exemplars that best capture the construct of interest rather than trying to generate a longer list of exemplars that are not high-quality representations of the category.

Presentation of single-discrimination trials (Stages 1, 2, and 5 in Figure 3.1) such that the target concepts precede the attribute traits allows the initial category construal to be uninfluenced by the subsequent attributes. Within each of the combined-task blocks (Stages 3, 4, 6, and 7), alternating stimuli from the attribute and trait categories provides participants a means of using the relevant features (category or attribute) for the task.

Study Design

Number of Trials

There are three main categorization tasks in the IAT: single-category classifications (Stages 1, 2, and 5 in Figure 3.1), one configuration of double categorizations (Stages 3 and 4), and an alternative configuration of double categorizations (Stages 6 and 7). The order of the two double configuration tasks is usually counterbalanced between subjects. Evidence indicates that including 20 trials in Stages 3 and 6 (the first sets of each combined pairing) and 40 trials in Stages 4 and 7 (the second sets of each combined pairing) yields good psychometric properties (Greenwald et al., 1998; Nosek et al., 2005) for the IAT, and it is not clear that there is any benefit to using more trials.

One of the most common artifacts observed on the IAT is the tendency for the first combined configuration to interfere with performance in the second combined configuration. For example, participants completing a gender-science stereotype IAT typically show larger IAT effects when the stereotype-consistent Male + Math block precedes the stereotype-inconsistent Female + Math block than vice versa. Nosek and colleagues (2005) varied the number of trials in Stage 4 of the IAT, during which the subject practices the single-category classification before beginning the second set of combined pairings. Use of 40 single-categorization trials at this stage reduced this undesirable order effect.

Order of Measures

Researchers face at least two decisions about counterbalancing measures when developing study designs. First, when the IAT is used as a predictor or criterion variable, in what order should measures be completed? Effects of the order of measures have been most widely considered as a potential moderator of the relationship between implicit and explicit measures. If completing explicit measures makes concepts more accessible (Fazio, 1995), then providing self-reports before the IAT may increase the extent to which the two measures tap a similar construct, thus inflating the correlation between them (see Bosson et al., 2000, for such a result). Contrary to this logic, however, Hofmann and colleagues' (2005) meta-analysis found that implicit–explicit correspondence did not differ when explicit measures preceded the IAT (52 independent observations, $\rho = .23$), as compared to when the IAT was completed first (48 independent observations, $\rho = .21$). Similarly, in an experimental investigation of this issue, Nosek and colleagues (2005) systematically varied the order of implicit and explicit measures of three IATs on a publicly available website and found that the relationship between implicit and explicit measures did not vary as a function of the order in which measures were completed (average $r = .23$ in both orders of presentation). Additional data from a large Internet sample ($N > 11,000$) indicated that across a larger number of attitude objects ($N = 57$), presentation order did not affect the relationship between implicit and explicit measures (Nosek, 2005).

These data appear to suggest that the order of implicit measures does not systematically affect the relationship between implicit and explicit measures. However, at least occasionally, the relationship between implicit and explicit measures does vary as a function of the order in which the IAT and explicit measures are completed (Bosson

et al., 2000). Design decisions are best made after giving careful thought to the appropriate order of measures.

The second question is related to the IAT's structure: Should the order of the combined conditions be counterbalanced? Because the well-documented order effects on the IAT may not always be eliminated even by additional practice trials in Stage 4 (Nosek et al., 2005), it is essential that researchers interested in the overall magnitude of the IAT effect counterbalance the presentation order of the combined pairings. Fixing the order may lead to an overestimate (if, for example, the flower + good stage in Figure 3.1 is always presented first) or underestimate (if, for example, the flower + bad stage in Figure 3.1 is always presented first) of the magnitude of the effect, as compared to other studies that counterbalance blocks.

Similarly, if the IAT is used as a predictor or criterion variable, the variability potentially added by counterbalancing block orders may make it more difficult for existing relationships to emerge. This intuition may suggest fixing the order of the blocks when interested in the predictive value of implicit cognitions. In practice, use of counterbalancing tends to have little effect on observed correlations with IAT measures. In Hofmann and colleagues' (2005) meta-analysis, correlations between the IAT and explicit measures were slightly higher when the presentation order of the combined pairings was counterbalanced (89 independent observations, $\rho = .25$) than when they were fixed (26 independent observations, $\rho = .18$). This difference is presumably due to decisions about counterbalancing having covaried with features of the task that at least mildly moderated implicit–explicit relationships, such as the attitude domain's social sensitivity. There has been no experimental analogue of this finding, although correspondence between implicit and explicit measures did not vary across the two possible task orders (Nosek et al., 2005). A researcher who wishes to account for the potential variability attributed to IAT order can include dummy variables that code for presentation order; this approach allows both estimation of mean IAT effects and accounts for variability due to counterbalancing.

Other Design Issues

The effects of a number of other features that can vary across IATs have been examined. By including error feedback (a red X that appears when an incorrect response is made) and requiring subjects to make the correct response before proceeding to the next trial, the researcher can ensure that items are categorized as intended. The amount of time between the response to a given stimulus and presen-

tation of the next stimulus is typically greater than 150 msec. Varying these intertrial intervals up to 750 msec did not affect the IAT results (Greenwald et al., 1998). In addition, a number of studies have ruled out differential familiarity as a plausible explanation for implicit biases as measured by the IAT (Dasgupta et al., 2003; Dasgupta, McGhee, Greenwald, & Banaji, 2000; Ottaway et al., 2001).

These findings represent knowledge based on large numbers of tests, but researchers are likely to encounter situations in which alterations may improve task performance. When time is a constraint, an investigator may choose to use fewer trials. When the task is likely to be especially challenging to a subject population, additional practice tasks may improve performance. Systematic variations of task features will likely lead to improvements of the task in general, or to development of specific variants of the task that are appropriate for particular contexts or populations; for example, systematic examination of task features led to the development of a child-friendly version of the IAT (Baron & Banaji, 2006; Dunham, Baron, & Banaji, 2006b).

DATA ANALYSIS AND INTERPRETATION

An Improved Scoring Algorithm

Until recently the majority of studies that included one or more IATs reported the IAT effect as the difference in mean (usually log-transformed) response latencies between the second of the two combined pairings (depicted as Stages 4 and 7 in Figure 3.1), with some adjustments for excessively slow or fast responses (see Greenwald et al., 1998). More recently, Greenwald and colleagues (2003), based on analyses of large data sets available from the public websites, developed an improved scoring method for IAT data. They identified the psychometrically best-functioning scoring procedure from a large number of candidate scoring methods. The recommended algorithm was the one that worked best to minimize (1) the correlation between IAT effects and individual subjects' average response latencies, (2) the effect of the order of the IAT blocks, and (3) the effect of previously completing one or more IATs on IAT scores, while (4) retaining strong internal consistency and (5) maximizing the correlation between implicit and explicit measures.

Based on these criteria, Greenwald and colleagues (2003) recommended the measure that they identified as *D* to replace the previously conventional scoring method. *D* is computed as the difference in average response latency between the IAT's two combined tasks

(e.g., flower + good, flower + bad), divided by an “inclusive” standard deviation of subject response latencies in the two combined tasks. Table 3.3 provides an overview of the specific stages of this approach.

Other Interpretation Issues

One of the structural features of the IAT is its relative nature—IAT effects are always a function of two target categories. Some researchers have used subsets of response latencies to each target category as an index for absolute attitudes. Nosek and colleagues (2005) examined the feasibility of this analysis strategy and concluded that such an approach is not appropriate. If absolute attitudes could be distilled from IAT data, they reasoned, such attitudes should correspond more highly with absolute explicit attitudes than relative explicit attitudes (Campbell & Fiske, 1959). In addition to calculating the standard (relative) IAT effect, they decomposed scores into two “absolute IAT” scores. Across four IATs, the separate components of the IAT, as well as the standard (relative) IAT score, had higher correspondence with relative, rather than absolute, self-reported attitudes. Responses to stimuli in the IAT are made in the framework of a comparison to a contrasting category, and this comparison is reflected in each trial response. Absolute implicit attitudes may be best assessed

TABLE 3.3. Summary of IAT Scoring Procedures Recommended by Greenwald et al. (2003)

-
- 1 Delete trials greater than 10,000 msec
 - 2 Delete subjects for whom more than 10% of trials have latency less than 300 msec
 - 3 Compute the “inclusive” standard deviation for all trials in Stages 3 and 6 and likewise for all trials in Stages 4 and 7
 - 4 Compute the mean latency for responses for each of Stages 3, 4, 6, and 7
 - 5 Compute the two mean differences ($\text{Mean}_{\text{Stage 6}} - \text{Mean}_{\text{Stage 3}}$) and ($\text{Mean}_{\text{Stage 7}} - \text{Mean}_{\text{Stage 4}}$)
 - 6 Divide each difference score by its associated “inclusive” standard deviation
 - 7 D = the equal-weight average of the two resulting ratios
-

Note. From Greenwald, Nosek, and Banaji (2003, Table 4). Copyright 2003 by the American Psychological Association. Adapted by permission. This computation is appropriate for designs in which subjects must correctly classify each item before the next stimulus appears. If subjects can proceed to the next stimulus following an incorrect response, the following steps may be taken between Steps 2 and 3 in the table: (1) compute mean latency of correct responses for each combined Stage (3, 4, 6, 7); (2) replace each error latency with an error penalty computed optionally as “Stage mean + 600 msec” or “Stage mean + twice the SD of correct responses for that Stage.” Proceed as above from Step 3 using these error-penalty latencies. Stage numbers refer to the stages depicted in Figure 3.1. SPSS and SAS syntax for implementing the new scoring algorithm are available at faculty.washington.edu/agg/iat_materials.htm and www.briannosek.com, respectively.

using an alternative implicit measure (De Houwer, 2003; Nosek & Banaji, 2001).

CONCLUSIONS

Because of the rapid dissemination of the IAT, researchers have correctly called for intensive investigation into its underlying psychometric properties and mechanisms. In the past few years investigations into these features have led to identification of confounding influences (Greenwald & Nosek, 2001; Greenwald et al., 2003; McFarland & Crouch, 2002; Mierke & Klauer, 2001), improvements to scoring strategies (Greenwald et al., 2003), and improvements in study designs (Nosek et al., 2005).

Nevertheless, a number of issues remain open and in critical need of analysis. A better understanding of the mechanism of the IAT is needed (Greenwald, Nosek, Banaji, & Klauer, 2005; Mierke & Klauer, 2003; Rothermund & Wentura, 2001, 2004). In addition, exploration of the relationship between changes in implicit cognitions and changes in behavior may help to identify mechanisms of behavioral change as well as consequences of the well-documented malleability effects. Rather than simply asking if the IAT converges with other implicit and explicit measures and covaries with meaningful criterion variables—because there is evidence that it does—the next generation of questions will likely continue the current shift to identifying when and why these patterns emerge. Answers to these questions will help in building theories of implicit social cognition, because methods are a central route to theory development.

ACKNOWLEDGMENTS

The research and writing of this chapter were supported by a grant from the National Institute of Mental Health and the Third Millennium Foundation as well as a fellowship from the Radcliffe Institute for Advanced Study to Mahzarin R. Banaji. We thank Dolly Chugh for her comments on a prior version of this chapter.

NOTE

1. Of course, we are not suggesting that the sample at the demonstration websites is random (see Nosek et al., 2002a, for a discussion of the benefits and challenges of conducting this kind of research over the Internet). The demonstration website (now at www.implicit.harvard.edu) has been operating continuously since Sep-

tember 1998. In addition, the Southern Poverty Law Center (SPLC) maintains a website devoted to educating visitors about bias in various forms. This website had included a component that, as at the demonstration site, allowed participants to select and complete one or more IATs and receive feedback on each test completed.

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.
- Akalis, S. A., Nannapaneni, J., & Banaji, M. R. (2006). *Do-it-yourself mental makeovers: Self-generated implicit attitude shifts in college students and yoga practitioners*. Unpublished manuscript.
- Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or would Jesse Jackson “fail” the Implicit Association Test? *Psychological Inquiry, 15*, 257–278.
- Asendorpf, J. B., Banse, R., & Muecke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83*, 380–393.
- Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. (2003). Black Americans’ implicit racial associations and their implications for intergroup judgment. *Social Cognition, 21*, 61–87.
- Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology, 81*, 789–799.
- Back, M., Schmukle, S. C., Egloff, B., & Gutenberg, J. (2005). Measuring task-switching ability in the Implicit Association Test. *Experimental Psychology, 52*, 167–179.
- Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.
- Banaji, M. R. (in press). Toward an understanding of the origins of attitudes. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new wave of implicit measures*. Mahwah, NJ: Erlbaum.
- Banaji, M. R., Bazerman, M. H., & Chugh, D. (2003). How (un)ethical are you? *Harvard Business Review, 30*, 1–20.
- Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes toward homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie, 48*, 145–160.
- Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6, 10, and adulthood. *Psychological Science, 17*, 53–58.
- Bassett, J. F., & Dabbs, J. M. (2003). Evaluating explicit and implicit death attitudes in funeral and university students. *Mortality, 8*, 352–371.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242–261.

- Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81*, 828–841.
- Bosson, J. K., Swann, W. B. J., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79*, 631–643.
- 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.
- Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology, 40*, 227–259.
- Bumiller, E. A. (2004, June 26). Bush gets chilly reception on eve of meeting in Ireland. *New York Times*, p. 7.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.
- Carney, D. R., Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (in press). Implicit Association Test (IAT). In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of social psychology*. Thousand Oaks, CA: Sage.
- Chassin, L., Presson, C., Rose, J., Sherman, S. J., & Probst, J. (2002). Parental smoking cessation and adolescent smoking. *Journal of Pediatric Psychology, 27*, 485–496.
- Citrin, L. B., & Greenwald, A. G. (1998). *Measuring implicit cognition: Psychologists' and entomologists' attitudes toward insects*. Paper presented at the Midwestern Psychological Association, Chicago.
- Conroy, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology, 89*, 469–487.
- Cunningham, W. A., Johnson, M. K., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2003). Neural components of social evaluation. *Journal of Personality and Social Psychology, 85*, 639–649.
- Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of Black and White faces. *Psychological Science, 15*, 806–813.
- Cunningham, W. A., Nezlek, J., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. *Personality and Social Psychology Bulletin, 30*, 1332–1346.
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12*, 163–170.
- Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin, 29*, 532–544.
- Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology, 40*, 642–658.

- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81*, 800–814.
- Dasgupta, N., Greenwald, A. G., & Banaji, M. R. (2003). The first ontological challenge to the IAT: Attitude or mere familiarity? *Psychological Inquiry, 14*, 238–243.
- Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology, 36*, 316–328.
- De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology, 37*, 443–451.
- De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology, 50*, 77–85.
- de Jong, P. J., van den Hout, M. A., Rietbroek, H., & Huijding, J. (2003). Dissociations between implicit and explicit attitudes toward phobic stimuli. *Cognition and Emotion, 17*, 521–545.
- Devos, T., & Banaji, M. R. (2005). American = White? *Journal of Personality and Social Psychology, 88*, 447–466.
- Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology, 86*, 345–355.
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica, 30*, 413–421. (Original work published 1868)
- Dunham, Y., Baron, A. S., & Banaji, M. R. (2006a). From American city to Japanese village: The omnipresence of implicit race attitudes. *Child Development, 77*, 1268–1281.
- Dunham, Y., Baron, A. S., & Banaji, M. R. (2006b). *The person and the group: A developmental analysis of consistency in implicit social cognition*. Unpublished manuscript.
- 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.
- Egloff, B., Schwerdtfeger, A., & Schmukle, S. C. (2005). Temporal stability of the Implicit Association Test. *Anxiety, 84*, 82–88.
- Fazio, R. H. (1995). Attitudes as object–evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247–282). Mahwah, NJ: Erlbaum.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013–1027.
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297–327.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Oxford, UK: Row, Peterson.
- Gabrieli, J. D. E., Fleischman, D. A., Keane, M. M., Reminger, S. L., & Morrell, F. (1995). Double dissociations between memory systems underlying explicit and implicit memory in the human brain. *Psychological Science, 6*, 76–82.

- Gawronski, B. (2002). What does the Implicit Association Test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology*, *49*, 171–180.
- Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression formation: Associations influence the construal of individuating information. *European Journal of Social Psychology*, *33*, 573–589.
- Gemar, M. C., Segal, Z. V., Sagrati, S., & Kennedy, S. J. (2001). Mood-induced changes on the Implicit Association Test in recovered depressed patients. *Journal of Abnormal Psychology*, *110*, 282–289.
- Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology*, *40*, 357–365.
- Gray, N. S., Brown, A. S., & MacCulloch, M. J. (2005). An implicit test of the associations between children and sex in pedophiles. *Journal of Abnormal Psychology*, *114*, 304–308.
- Gray, N. S., MacCulloch, M. J., Smith, J., Morris, M., & Snowden, R. J. (2003). Violence viewed by psychopathic murderers. *Nature*, *423*, 497–498.
- Green, A., Carney, D. R., Pallin, D., Iezzoni, L., & Banaji, M. R. (2006). *Physicians' implicit biases predict differential treatment of Black versus White patients*. Unpublished manuscript.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.
- 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. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie*, *48*, 85–93.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.
- Greenwald, A. G., Nosek, B. A., Banaji, M. R., & Klauer, K. C. (2005). Validity of the salience asymmetry interpretation of the IAT: Comment on Rothermund and Wentura, 2004. *Journal of Experimental Psychology: General*, *134*, 420–425.
- Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*, 1–20.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- 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.

- Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science, 14*, 640–643.
- Hugenberg, K., & Bodenhausen, G. V. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in race. *Psychological Science, 15*, 342–345.
- Hummert, M. L., Garstka, T. A., O'Brien, L. T., Greenwald, A. G., & Mellott, D. S. (2002). Using the Implicit Association Test to measure age differences in implicit social cognitions. *Psychology and Aging, 17*, 482–495.
- Jajodia, A., & Earleywine, M. (2003). Measuring alcohol expectancies with the Implicit Association Test. *Psychology of Addictive Behaviors, 17*, 126–133.
- Jellison, W. A., McConnell, A. R., & Gabriel, S. (2004). Implicit and explicit measures of sexual orientation attitudes: Ingroup preferences and related behaviors and beliefs among gay and straight men. *Personality and Social Psychology Bulletin, 30*, 629–642.
- Jordan, C. H., Spencer, S. J., Zanna, M. P., Hoshino-Browne, E., & Correll, J. (2003). Secure and defensive high self-esteem. *Journal of Personality and Social Psychology, 85*, 969–978.
- Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology, 33*, 1–27.
- Jost, J. T., Banaji, M. R., & Nosek, B. A. (2005). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology, 25*, 881–919.
- Jost, J. T., Pelham, B. W., & Carvallo, M. R. (2002). Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups. *Journal of Experimental Social Psychology, 38*, 586–602.
- Kang, J., & Banaji, M. R. (2006). Fair measures: A behavioral realist revision of “affirmative action.” *California Law Review*, pp. 1063–1118.
- Karpinski, A. (2004). Measuring self-esteem using the Implicit Association Test: The role of the other. *Personality and Social Psychology Bulletin, 30*, 22–34.
- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 774–788.
- Karpinski, A., Steinman, R. B., & Hilton, J. L. (2005). Attitude importance as a moderator of the relationship between implicit and explicit attitude measures. *Personality and Social Psychology Bulletin, 31*, 949–962.
- 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.
- Kraus, S. J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin, 21*, 58–75.
- Kuhnen, U., Schiessl, M., Bauer, N., Paulig, N., Pohlmann, C., & Schmidhals, K. (2001). How robust is the IAT? Measuring and manipulating implicit attitudes of East- and West-Germans. *Zeitschrift für Experimentelle Psychologie, 48*, 135–144.
- Lane, K. A., & Banaji, M. R. (2004). *Implicit intergroup bias: The contributions of*

- ingroup liking and outgroup disliking*. Paper presented at the 5th annual meeting of the Society for Personality and Social Psychology, Austin, TX.
- Lane, K. A., Mitchell, J. P., & Banaji, M. R. (2005). Me and my group: Cultural status can disrupt cognitive consistency. *Social Cognition, 23*, 353–386.
- LaPierre, R. (1934). Attitudes and actions. *Social Forces, 13*, 230–237.
- Livingston, R. W. (2002). The role of perceived negativity in the moderation of African Americans' implicit and explicit racial attitudes. *Journal of Experimental Social Psychology, 38*, 405–413.
- Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81*, 842–855.
- Maison, D., Greenwald, A. G., & Bruin, R. (2001). The Implicit Association Test as a measure of implicit consumer attitudes. *Polish Psychological Bulletin, 32*, 61–69.
- Marsh, K. L., Johnson, B. T., & Scott-Sheldon, L. A. J. (2001). Heart versus reason in condom use: Implicit versus explicit attitudinal predictors of sexual behavior. *Zeitschrift für Experimentelle Psychologie, 48*, 161–175.
- McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination and racism* (pp. 91–126). San Diego, CA: Academic Press.
- McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37*, 435–442.
- McFarland, S. G., & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition, 20*, 483–510.
- Mierke, J., & Klauer, K. C. (2001). Implicit association measurement with the IAT: Evidence for effects of executive control processes. *Zeitschrift für Experimentelle Psychologie, 48*, 107–122.
- Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology, 85*, 1180–1192.
- Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General, 132*, 455–469.
- Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology, 83*, 1029–1050.
- Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition, 19*, 395–417.
- Neumann, R., Hulsenbeck, K., & Seibt, B. (2004). Attitudes towards people with AIDS and avoidance behavior: Automatic and reflective bases of behavior. *Journal of Experimental Social Psychology, 40*, 543–550.
- Nock, M., & Banaji, M. R. (2006). *Assessing suicide risk implicitly*. Unpublished manuscript.
- 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*, 625–666.
- Nosek, B. A., & Banaji, M. R. (2002). (At least) two factors moderate the relationship between implicit and explicit attitudes. In R. K. Ohme & M. Jarymowicz (Eds.), *Natura Automatyzmów* (pp. 49–56). Warsaw: WIP PAN & SWPS.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002a). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics, 6*, 101–115.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002b). Math = male, me = female, therefore math is not equal to me. *Journal of Personality and Social Psychology, 83*, 44–59.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test II: Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*, 166–180.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (in press). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Automatic processes in social thinking and behavior*. Hove, UK: Psychology Press.
- Nosek, B. A., & Hansen, J. J. (2004). *The associations in our head belong to us: Searching for attitudes and knowledge in implicit cognition*. Unpublished manuscript.
- Nosek, B. A., & Smyth, F. L. (in press). A multitrait–multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology*.
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Linder, N., Smith, C. T., et al. (2006). *Pervasiveness and variability in implicit bias*. Unpublished manuscript.
- Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science, 14*, 636–639.
- Olson, M. A., & Fazio, R. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653–667.
- Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (2001). Implicit attitudes and racism: Effects of word familiarity and frequency on the Implicit Association Test. *Social Cognition, 19*, 97–144.
- Palfai, T. P., & Ostafin, B. D. (2003). Alcohol-related motivational tendencies in hazardous drinkers: Assessing implicit response tendencies using the modified-IAT. *Behaviour Research and Therapy, 41*, 1149–1162.
- Perruchet, P., & Baveux, P. (1989). Correlational analyses of explicit and implicit memory performance. *Memory and Cognition, 17*, 77–86.
- Perugini, M. (2005). Predictive models of implicit and explicit attitudes. *British Journal of Social Psychology, 44*, 29–45.
- Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology, 48*, 609–647.
- Petty, R. E., Wheeler, S. C., & Tormala, Z. L. (2003). Persuasion and attitude change. In T. Millon & M. J. Lerner (Eds.), *Handbook of psychology: Personality and social psychology* (Vol. 5, pp. 353–382). Hoboken, NJ: Wiley.
- Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J.

- C., Gore, J. C., et al. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, *12*, 729–738.
- Pinter, B., & Greenwald, A. G. (2005). Clarifying the role of the “other” category in the self-esteem IAT. *Experimental Psychology*, *52*, 74–79.
- Poehlman, A., Uhlmann, E., Greenwald, A. G., & Banaji, M. R. (2005). *Understanding and using the Implicit Association Test III: A meta-analysis of predictive validity*. Unpublished manuscript.
- Richeson, J. A., & Ambady, N. (2003). Effects of situational power on automatic racial prejudice. *Journal of Experimental Social Psychology*, *39*, 177–183.
- Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., et al. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nature Neuroscience*, *6*, 1323–1328.
- Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, *14*, 287–290.
- Roediger, H. L. III. (1990). Implicit memory: Retention without remembering. *American Psychologist*, *45*, 1043–1056.
- Roediger, H. L. III. (2003). Reconsidering implicit memory. In J. S. Bowers & C. J. Marsolek (Eds.), *Rethinking implicit memory* (pp. 3–18). New York: Oxford University Press.
- Rosenberg, M. (1989). *Society and the adolescent self-image* (Rev. ed.). Middletown, CT: Wesleyan University Press.
- Rothermund, K., & Wentura, D. (2001). Figure–ground asymmetries in the Implicit Association Test (IAT). *Zeitschrift für Experimentelle Psychologie*, *48*, 94–106.
- 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., Feinberg, J., & Fairchild, K. (2002). Minority members’ implicit attitudes: Automatic ingroup bias as a function of group status. *Social Cognition*, *20*, 294–320.
- 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.
- Rudman, L. A., & Heppen, J. B. (2003). Implicit romantic fantasies and women’s interest in personal power: A glass slipper effect? *Personality and Social Psychology Bulletin*, *29*, 1357–1370.
- Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin*, *26*, 1315–1328.
- Rudman, L. A., & Lee, M. R. (2002). Implicit and explicit consequences of exposure to violent and misogynous rap music. *Group Processes and Intergroup Relations*, *5*, 133–150.
- Schmukle, S. C., & Egloff, B. (2004). Does the Implicit Association Test for assessing anxiety measure trait and state variance? *European Journal of Personality*, *18*, 483–494.

- Schultz, P. W., Shriver, C., Tabanico, J. J., & Khazian, A. M. (2004). Implicit connections with nature. *Journal of Environmental Psychology, 24*, 31–42.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93–105.
- Sherman, S. J., Rose, J. S., Koch, K., Presson, C. C., & Chassin, L. (2003). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology, 22*, 13–39.
- Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 108–131.
- Steffens, M. C., & Buchner, A. (2003). Implicit Association Test: Separating transsituationally stable and variable components of attitudes toward gay men. *Experimental Psychology, 50*, 33–48.
- Steffens, M. C., & Plewe, I. (2001). Items' cross-category associations as a confounding factor in the Implicit Association Test. *Zeitschrift für Experimentelle Psychologie, 48*, 123–134.
- Swanson, J. E., Rudman, L. A., & Greenwald, A. G. (2001). Using the Implicit Association Test to investigate attitude-behaviour consistency for stigmatised behaviour. *Cognition and Emotion, 15*, 207–230.
- Teachman, B. A., Gregg, A. P., & Woody, S. R. (2001). Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology, 110*, 226–235.
- Teachman, B. A., & Woody, S. R. (2003). Automatic processing in spider phobia: Implicit fear associations over the course of treatment. *Journal of Abnormal Psychology, 112*, 100–109.
- Uhlmann, E., & Swanson, J. (2004). Exposure to violent video games increases automatic aggressiveness. *Journal of Adolescence, 27*, 41–52.
- Wiers, R. W., van Woerden, N., Smulders, F. T. Y., & de Jong, P. J. (2002). Implicit and explicit alcohol-related cognitions in heavy and light drinkers. *Journal of Abnormal Psychology, 111*, 648–658.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101–126.
- Yamaguchi, S., Greenwald, A. G., Banaji, M. R., Murakami, F., Chen, D., Shiomura, K., et al. (2006). *Comparisons of implicit and explicit self-esteem among Chinese, Japanese, and North American university students*. Unpublished manuscript.