numerical algorithm
ı model parameters.
ıgy, 52, 1–9.
coding and memory.
gy, 94, 90–100.
ı, directionality, and
Experimental Psychol-

ffective and associa-
al decision task: Yes
ts reveal evaluative
Experimental Psychol-
ıition, 26, 456–469.
(2003). Age stereo-
ıen: Analyses of ac-
tence-priming task.
26.
. Automatic evalua-
of masked affective
ition and Emotion.
epeated masked cat-
ed exemplars: New
priming. Journal of
, Memory, and Cog-

). Response-bound
ng in the naming
74–384.
/ho wants to be . . .
automatic adapta-
nition, 22, 30–53.
ssessing the struc-
self-defensive pro-
task. Self and Iden-

W. (2005). Masked
Evidence for a cor-
t self-esteem. Jour-
y, 41, 654–663.
003). The "med-
A general model
J. Musch & K. C.
aluation: Affective
pp. 51–86). Mah-

riming semantic?
ology, 8, 113–161.
ıttitudes through
Schwarz (Eds.),
7–58). New York:

, B. (1997). Evi-
licit level and its
asures. Journal of
262–274.
B. (2001). Evalu-
n automatic ste-
Experimental So-

CHAPTER 7

A Practical Guide to Implicit Association Tests and Related Tasks

Sarah Teige-Mocigemba, Karl Christoph Klauer, and Jeffrey W. Sherman

The story of the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is one of an incredible boom. Eleven years after its first publication, more than 450 articles have been published that either applied the method to the assessment of various "implicit" constructs (for a definition of the term *implicit* and its different uses, see De Houwer, 2006; De Houwer & Moors, 2007; De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009) or investigated the underlying processes and possible confounds of IAT effects. In the present chapter, we trace this story of the IAT by reviewing research on the IAT and related tasks. In particular, we initially identify factors that contributed to the explosion of IAT research. We then introduce the IAT methodology, review findings of its psychometric properties, and present process models that have been proposed to account for IAT effects. In the course of discussing research on contaminations of IAT effects, we finally introduce IAT-related tasks, most of which have been developed as possible solutions to IAT confounds. Other possible remedies for the respective confounds are also presented.

ON A NEW APPROACH TO AN OLD PROBLEM

Since researchers aim at assessing core psychological processes, they are faced with two key problems of direct measures (e.g., self-reports), namely introspective limits (e.g., Nisbett & Wilson, 1977) and susceptibility to self-presentation or socially desirable responding (e.g., Paulhus, 1984). The idea that there might be more about ourselves than we can tell or want to tell promoted the development of several indirect measures. Such measures were developed in the hope of obtaining diagnostic inferences about a person's dispositions without having to ask the person directly. This hope, however, was soon dampened as findings of unacceptable reliability and validity questioned the usefulness of early indirect measures such as projective tests (e.g., Lilienfeld, Wood, & Garb, 2000).

With technological progress making it possible to present stimuli and record response times with highly accurate computer-based methods, a new class of indirect measures emerged, namely response time measures (also termed "implicit" measures).[1] Assuming that response time patterns may reflect the associative processes of interest appropriately (see Strack & Deutsch, 2004), such measures are expected to offer straightforward access to cognitive structures or processes. Researchers applied experimental paradigms of cognitive psychology such as sequential priming (Neely, 1977) or response interference tasks (Kornblum, Hasbroucq, & Osman, 1990) to the assessment of attitudes, stereotypes, self-esteem, and personality traits (for reviews, see Fazio & Olson, 2003; Schnabel, Asendorpf, & Greenwald, 2008; Wittenbrink & Schwarz, 2007; for similarities of and differences between response time measures, see De Houwer, 2001, 2003b, 2008). For the new computer-based measures, the known problems of early indirect measures initially seemed to recur: The new measures proved to be useful tools to examine differences at the group level, but scarcely any of them met the test-theoretical criteria required for the assessment of differences at the individual level. Thus, like for the early indirect measures, most implicit response time measures suffered from low to, at best, moderate reliability. In 1995, Greenwald and Banaji argued that because such unreliable measures fail to detect interindividual differences, their application to the assessment of implicit constructs at the individual level is highly problematic. Three years later, the IAT was introduced as the first implicit response time measure that proved to be reliable, at least in terms of internal consistency (see later discussion). Thus, the great demand for *reliable* indirect measures helped to make the IAT widely accepted. The IAT's easy applicability and effective promotion might also have contributed to its popularity and widespread use in diverse subdisciplines of psychological research. In the following section, we give more detailed information about the general procedure, implementation details, and different scoring procedures of the IAT. We then turn to the IAT's psychometric properties.

THE IAT

Procedure

The IAT is thought to assess the strength of associations between target categories (e.g., black persons vs. white persons) and attribute categories (e.g., negative vs. positive), both arranged on bipolar dimensions, by comparing the response latencies for two differently combined categorization tasks. Participants are instructed to categorize stimuli that represent the four categories (e.g., names typical for blacks vs. whites and negative vs. positive words) with the help of two response keys, each assigned to two of the four categories. The IAT's basic assumption is that if two concepts are highly associated, categorization will be easier when the two associated categories share the same response (in the so-called *compatible* block; De Houwer, 2003b) than when they require different responses (in the so-called *incompatible* block; De Houwer, 2003b).

Table 7.1 presents a typical task sequence of the IAT (here, a racial attitude IAT) consisting of seven blocks, some of which are practice blocks to acquaint participants with the stimulus material and categorization rules. In a racial attitude IAT,

TABLE 7.1. Example of a Racial Attitude Implicit Association Test (IAT): Task Sequence

Block	*N* trials	Task	Response key assignment	
			Left key	Right key
1	20	Target discrimination	Black	White
2	20	Attribute discrimination	Negative	Positive
3	20	Initial combined task	Black, negative	White, positive
4	40	Initial combined task	Black, negative	White, positive
5	20 or 40	Reversed target discrimination	White	Black
6	20	Reversed combined task	White, negative	Black, positive
7	40	Reversed combined task	White, negative	Black, positive

Note. Nosek, Greenwald, and Banaji (2005) recommend increasing the number of trials in the fifth block as an effective means to reduce compatibility-order effects (see Confounding Factors of the IAT Effect section). Therefore, some IAT procedures present 40 instead of 20 trials in this block, in which participants practice the reverse target discrimination task.

for instance, participants are trained to press a left key for "black" stimuli and a right key for "white" stimuli in the first block of 20 trials (target discrimination). In the second block of 20 trials, they are trained to press the same left key for "negative" stimuli and the same right key for "positive" stimuli (attribute discrimination). The diagnostically relevant third block (20 trials) and fourth block (40 trials) combine the attribute and target discrimination. Participants now are to respond left to negative and black stimuli and right to positive and white stimuli. In the fifth block (20 or 40 trials; see later discussion), target discrimination is reversed: Participants are trained to press the left key for white stimuli and the right key for black stimuli. The diagnostically relevant sixth block (20 trials) and seventh block (40 trials) again combine the attribute and the previously reversed target discrimination. Participants now are to respond left to negative and white stimuli, and right to positive and black stimuli. The difference in performance between the initial combined blocks (third and fourth blocks) and the reversed combined blocks (sixth and seventh blocks) is called the IAT effect. Subdivision of the initial combined and reversed combined categorization tasks into two blocks of different lengths (20 vs. 40 trials), respectively, has mainly historical reasons: In contrast to current scoring procedures (see later discussion) that include data from both the shorter combined blocks (third and sixth blocks) and the longer combined blocks (fourth and seventh blocks) in the algorithms (Greenwald, Nosek, & Banaji, 2003), early scoring procedures denoted the third and the sixth blocks as practice blocks and excluded them from the analyses (Greenwald et al., 1998).

The IAT effect is interpreted in its size and direction as revealing the relative association strength between the target and attribute categories. Accordingly, individuals with implicit prejudices against blacks are expected to respond faster and more accurately when black stimuli and negative attributes are assigned to the same response key (and white stimuli and positive attributes to the other key) compared with the reverse configuration (black and positive are assigned to one key, white and negative to the other key). It is important to note that IAT effects always have to be interpreted in a relative manner (for more details, see Relative Measure section): A racial attitude IAT effect, for instance, does not permit any conclusions about an individual's evaluation of blacks but provides only information about an individual's preference for blacks over whites (or whites over blacks).

Implementation Details

IATs as well as other response time–based measures can be easily implemented using software packages such as Inquisit, E-Prime, Direct-RT, and SuperLab to name just a few (see Stahl, 2006, for a comparison of these four software packages). Sample programs can be downloaded from several websites (e.g., *faculty.washington.edu/agg*).

Different procedural details have been used in IAT implementations. In most instances, procedural variations did not considerably affect IAT effects or their correlations (see Nosek, Greenwald, & Banaji, 2007). Those procedural variations that have been shown to have an impact on IAT effects are discussed later (see Confounding Factors of IAT Effects section). Implementation details of current standard IAT procedures comprise (1) the instruction to respond as quickly and accurately as possible, (2) correction of erroneous responses as indicated by an error cue (e.g., presentation of a red X or the word *error* below the stimulus), (3) display of category labels assigned to the left or right response key in the corresponding upper screen corners throughout all blocks, (4) intertrial interval of 150 msec–750 msec (250 msec may be most often used), (5) five to six stimuli per category (at least two), (6) alternation between target and attribute stimuli in the combined blocks, (7) otherwise randomized trial order if group differences are the main focus of the experiment, but (8) a fixed random trial order for all participants in correlational studies in order to reduce confounds of procedural and interindividual variance (e.g., Banse, Seise, & Zerbes, 2001).

Scoring Procedures

Different scoring procedures have been proposed for calculating IAT effects (see Greenwald et al., 2003). All scoring procedures compare the performance between the initial combined blocks (see Table 7.1, third and fourth blocks) and the reverse combined blocks (see Table 7.1, sixth and seventh blocks). Because of the variety of available IAT score calculations and different recommendations about which score should be used, most researchers report analyses based on both the so-called conventional IAT score and the so-called *D* measures. According to the conventional algorithm (see Greenwald et al., 2003), the performance difference between the two combined tasks is based on log-transformed response latencies, with latencies smaller than 300 msec or greater than 3,000 msec being recoded to 300 msec and 3,000 msec, respectively. For descriptive statistics, however, IAT scores based on raw response latencies are often reported.

Greenwald and colleagues (2003) suggested improved scoring procedures for the IAT, the *D* measures, which were optimized with regard to the IAT's psychometric criteria (e.g., increased internal consistency, higher correlations with explicit measures, resistance to some extraneous procedural influences). *D* measures differ from the conventional algorithm in several aspects, including modified upper and lower tail treatment of latencies, inclusion of both correct and incorrect responses, with incorrect response latencies being increased by an error penalty, and an individual standardization similar to that in Cohen's effect size measure *d* (see Greenwald et al., 2003; Nosek et al., 2007). SPSS syntaxes for different *D* measures can be downloaded from Greenwald's website (*faculty.washington.edu/agg/iat_materials.htm*). It should be noted that although *D* measures are widely used by now, there is ongoing debate regarding the adequacy of the criteria according to which the algorithms of *D* measures have been developed (e.g., Wentura & Rothermund, 2007). In particular, researchers have criticized that *D* measures have been selected by maximizing the IAT's correlations with explicit measures. This might make the measure more direct on an indirect–direct dimension, which contravenes the basic idea of developing implicit measures. Clearly, more research is needed to evaluate strengths and weaknesses of the *D* measures compared with the conventional scores.

Regardless of the particular chosen algorithm, the practice of reducing eight distinct response latencies or error rates (for two targets and two attributes in both compatible and incompatible blocks) to a single number representing an IAT effect may have significant shortcomings. Difference scores may conceal important information about which particular trials are responsible for differences across conditions. Likewise, when reduced to a single index, it is impossible to examine whether different IAT scores in different experimental conditions, for example, reflect differences in responses to target trials, attribute trials, or both. Similarly, it is impossible to tell whether the effects are driven by responses on compatible trials, incompatible trials, or both. Important empirical and theoretical questions may be addressed by examining response latencies and error rates for each of the conditions separately (e.g., Brendl, Markman, & Messner, 2001).

Finally, recent research has suggested the use of more complex mathematical procedures such as diffusion model analysis (Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007) or multinomial modeling (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005) to analyze IAT data. Such mathematical models acknowledge that hardly any measure used in psychology is process pure in the sense that its outcome covaries only with differences in the construct that is to be measured. Thus, the modeling approach attempts to quantify the relative contributions of qualitatively different processes within a given measure (for a review, see Sherman, Klauer, & Allen, Chapter 9, this volume). To what extent the modeling approach will become accepted as a standard tool for analyzing the outcome of response time measures such as the IAT will probably depend on (1) accessibility of manageable software applications (e.g., Stahl & Klauer, 2007; Voss & Voss, 2007), (2) superiority to the conventional and/or *D* measures (e.g., regarding confounds), and (3) psychometric properties of the process components.

PSYCHOMETRIC PROPERTIES OF THE IAT

Reliability

Much of the IAT's popularity may be attributable to its comparatively satisfactory reliability estimates. In particular, internal consistencies (split-half correlations or Cronbach's alpha) have been shown to be satisfactory, with scores ranging from .70 to .90 (Nosek et al., 2007), differing slightly depending on the method of calculation (see Schnabel et al., 2008). By contrast, test–retest reliability has been found to be less satisfactory, ranging from .25 to .69 with mean and median estimates of about .50 (Lane, Banaji, Nosek, & Greenwald, 2007), varying little with retest interval (Egloff, Schwerdtfeger, & Schmukle, 2005). Explaining the discrepancy between the IAT's satisfactory internal consistency on the one hand and its lower test–retest reliability on the other is still an unresolved puzzle regarding the IAT's ability to capture temporally stable implicit constructs (e.g., personality traits or stable attitudes). Researchers have put forward different explanations for this discrepancy.

First, it has been argued that the IAT might measure states rather than traits, as suggested by studies showing the IAT's sensitivity to context effects and experimental manipulations (for reviews, see Blair, 2002; Gawronski & Bodenhausen, 2006). Drawing on such findings, some researchers have even questioned the existence of invariant, trait-like cognitive structures (e.g., Schwarz, 2007).

Second, it has been hypothesized that additional sources of construct-unrelated variance might lower test–retest reliability. For instance, the two measurement occasions may exert different influences on the processes underlying the IAT (not on the to-be-measured construct itself!) and thus lead to changes in the extent to which the construct in question causes variation in the IAT outcome (e.g., De Houwer, 2008; Gawronski, Deutsch, LeBel, & Peters, 2008). Such processing differences may comprise differences in (1) test-taking strategies (Egloff et al., 2005), (2) attentional foci when completing IATs (Gawronski et al., 2008), (3) learning effects (Schmukle & Egloff, 2004), or (4) other component response processes that do not reflect associations per se (e.g., Sherman et al., 2008).

Context effects may thus rely on (1) the IAT's sensitivity to changes in the construct of interest (i.e., IAT as a state measure), (2) its susceptibility to changes in additional sources of construct-unrelated variance (i.e., IAT as an insufficiently reliable trait measure), or (3) both (i.e., IAT as capturing both state- and trait-specific variation; Schmukle & Egloff, 2005; Sherman et al., 2008). In this regard, Gschwendner, Hofmann, and Schmitt (2008) recently emphasized the impact of construct accessibility on the IAT's temporal stability. They showed that (1) the IAT's test–retest reliability was enhanced in situations in which contextual background features activated specific construct-relevant concepts and that (2) this effect was particularly pronounced for individuals with chronically high accessibility for the relevant concept. These findings suggest that the IAT's ability to assess traits is enhanced by activating more specific representations via context information. Such representations may be assumed to reflect interindividually different, temporally stable patterns of associative activation (see also Conrey & Smith, 2007).

Validity

Group Level

UNIVERSAL ATTITUDES

At the group level, a priori assumptions have been used to examine the IAT's validity. For example, normative studies and a priori arguments suggest that there are objects toward which most people have relatively uniform attitudes (e.g., most people prefer flowers over insects). Accordingly, such universal attitudes should be reflected in the IAT effect. Indeed, flower–insect IATs have repeatedly been found to show more positive attitudes toward flowers than toward insects (for the first demonstration of this effect, see Greenwald et al., 1998) indicating validity of the IAT.

KNOWN-GROUP APPROACH

The so-called known-group approach contrasts groups that are assumed a priori to differ regarding the construct of interest. For some domains, the IAT proved to be valid as it revealed such differences (see Greenwald & Nosek, 2001). For instance, white and black individuals differed in their racial attitude IAT effects (Nosek, Banaji, & Greenwald, 2002), and a homosexuality attitude IAT distinguished between homosexuals and heterosexuals (Banse et al., 2001). In other domains, particularly those related to addictive behavior, the IAT did not consistently differentiate between groups (such as smokers vs. nonsmokers; Swanson, Rudman, & Greenwald, 2001; but see Perugini, 2005). The strength of the known-group approach, of course, hinges on the certainty with which the groups (e.g., smokers and nonsmokers) can be assumed to differ on a priori grounds.

EXPERIMENTALLY MANIPULATED ATTITUDES

Assuming that the IAT effect reflects the construct in question, experimental manipulation of this construct should influence the IAT effect in the expected manner. Olson and Fazio (2001), for instance, drew on this assumption and found evidence for the IAT's validity. Novel attitudes were formed by pairing previously unknown stimuli with other, clearly positive or negative stimuli. Results showed that IAT effects reflected these new attitudes, even when participants were unaware of its origins.

CRITICISM OF VALIDATION APPROACHES AT THE GROUP LEVEL

Importantly, the experimental validation approach has its limits when constructs are to be assessed that are expected to be stable over time. Per definition, such stable constructs (e.g., the personality trait of anxiousness) should not be affected by short-term manipulations. For instance, Schmukle and Egloff (2004) experimentally induced state anxiety by a public speaking task and found no effects on anxiety IATs. They interpreted their findings not in terms of the IAT's invalidity, but rather in terms of the IAT's validity as a measure of *trait* anxiety (see state–trait issue discussed previously). Similarly, the conclusiveness of the validation approaches by universal attitudes and known groups is limited: It cannot be ruled out that there are other, uncontrolled variables confounded with universal attitudes (e.g., stimulus selection) or group membership (e.g., cognitive abilities) that also account for the findings.

Individual Level

Most research on the IAT's validity adopted the correlational approach. Specifically, the IAT's validity has been investigated in terms of its correlations with (1) explicit measures and (2) other implicit measures and (3) by its predictive validity for behavioral measures.

CORRELATIONS WITH EXPLICIT MEASURES

Most studies have concentrated on implicit–explicit consistency. A meta-analysis over various content domains (including attitudes, stereotypes, and self-concept) revealed a somewhat low correlation of .24 between IATs and explicit measures (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005), whereas a large-scale analysis of Internet data from the IAT website yielded a higher implicit–explicit correlation of .37 (Nosek, 2005). Nosek's analysis might reveal somewhat higher correlations because, first, his data refer to attitude domains for which higher implicit–explicit consistency is expected. Second, Nosek exclusively used relative thermometer scales as explicit measures, which may better correspond to the IAT in that they are relative measures and more directly tap into an affective component. Third, greater variability in Nosek's Internet data might also have contributed to the higher implicit–explicit correlations.

There is still considerable controversy, however, as to whether such low to moderate correlations between the IAT and explicit measures should be interpreted as indices of discriminant validity or convergent validity (Payne, Burkley, & Stokes, 2008; see also Nosek & Smyth, 2007). The core of this debate traces back to the question of differences and similarities of the cognitive structures that underlie implicit and explicit measures. Some researchers postulate independent representations of implicit versus explicit constructs and thus interpret implicit–explicit correlations as indices of discriminant validity (e.g., Wilson, Lindsey, & Schooler, 2000). Other researchers postulate only one representation that can be tapped differently (i.e., using implicit or explicit measures) and, consequently, interpret implicit–explicit correlations as indices of convergent validity (e.g., Fazio, 1990; Nier, 2005).

It is to be noted, however, that recent research advised caution in interpreting implicit–explicit correlations as evidence for underlying cognitive structures. For instance, Payne and colleagues (2008) argued that measures differ with regard to several (structural) features and showed that structural fit has a strong impact on implicit–explicit correlations: The more similar the task demands of implicit and explicit measures, the higher the correlation between them, even when controlling for common method-specific variance. Implicit–explicit correlations might, therefore, reflect (structural) fit, or misfit, of the underlying measures rather than of the underlying cognitive structures (see also Hofmann, Gschwendner, Nosek, & Schmitt, 2005).

CORRELATIONS WITH OTHER IMPLICIT MEASURES

Considering the interpretation problems of implicit–explicit correlations, it has been suggested to focus on implicit–implicit correlations. Assuming that implicit measures capture the same (i.e., implicit) construct, the IAT's correlations with other implicit measures should reflect the IAT's convergent validity (e.g., Banaji, 2001). Interestingly, correlations between IATs and other implicit measures have typically been found to be weak (e.g., Olson & Fazio, 2003; Rudolph, Schröder-Abé, Schütz, Gregg, & Sedikides, 2008; Sherman, Presson, Chassin, Rose, & Koch, 2003; Teige, Schnabel, Banse, & Asendorpf, 2004). Low implicit–implicit consistency, however, is often not attributed to the IAT's invalidity but rather to two other factors. First, implicit measures other than the IAT often show unacceptable reliability estimates (Nosek et al., 2007). Because reliability sets upper limits on the to-be-expected correlation, implicit–implicit relations might necessarily be underestimated (Teige et al., 2004; see Cunningham, Preacher, & Banaji, 2001, for an approach to correct for such measurement error through latent variable analysis).

Second, not only implicit–explicit consistency but also implicit–implicit consistency might be influenced by the structural fit of the measures (Payne et al., 2008; see also De Houwer, 2008). Empirical evidence for this assumption is provided by studies that approximated formerly dissimilar features of implicit measures and indeed found higher implicit–implicit correlations. For example, Olson and Fazio (2003) argued that, as a result of different task demands of the IAT and affective priming, the IAT reveals evaluations of superordinate categories, whereas affective priming reveals evaluations of specific category exemplars used as stimuli. When affective priming was made more similar to the IAT by encouraging the primes' categorization in terms of the superordinate category, the correlation between affective priming and IAT was increased. In a similar vein, Steffens, Kirschbaum, and Glados (2008) equated the IAT and a response-window priming task with regard to stimulus selection (i.e., both tasks used only the concept categories as stimuli). Again, IAT effects and priming effects correlated significantly. Taken together, the IAT's low correlations with other implicit measures do not necessarily indicate the IAT's invalidity but can be accounted for by (1) low reliability estimates of implicit measures other than the IAT and (2) structural differences among the measures (Rudolph et al., 2008).

PREDICTIVE VALIDITY FOR BEHAVIORAL MEASURES

Most convincing in light of these discussions are correlational studies that have demonstrated the IAT's ability to predict behavior over and above explicit measures. Perugini (2005) distinguished between three different models of predictive validity of implicit and explicit measures: the additive, the multiplicative, and the double-dissociation model. Research on the IAT found evidence for all three models. As proposed by the additive model, the IAT and explicit measures explained separate portions of relevant criterion variance (e.g., Schnabel, Banse, & Asendorpf, 2006a). As suggested by the multiplicative model, the IAT and explicit measures interacted in predicting relevant behavior (e.g., Schröder-Abé, Rudolph, & Schütz, 2007). Finally, as proposed by the double-dissociation model, only the IAT predicted spontaneous behavior, whereas only explicit measures predicted controlled behavior (e.g., Asendorpf, Banse, & Mücke, 2002).

Evidence for the predictive validity of IATs across various behavioral domains is also provided by a recent meta-analysis by Greenwald, Poehlman, Uhlmann, and Banaji (2009). In socially sensitive domains such as stereotyping and prejudice, the IAT showed better predictive validity than explicit measures. This might have been expected, given that, particularly in these domains, socially desirable responding may bias explicit measures. In contrast, the meta-analysis revealed lower predictive validity for IATs than for explicit measures in studies that explored brand preferences or political attitudes. Importantly, in domains related to health behavior, the IAT has been shown to have weaker predictive validity: An IAT designed to assess preferences for apples versus candy bars did not predict the subsequent choice between an apple and a candy bar (e.g., Karpinski & Hilton, 2001; Spruyt, Hermans, De Houwer, Vandekerckhove, & Eelen, 2007). The IAT's insufficiency in such domains has been argued to be due to its sensitivity to so-called "extrapersonal" knowledge (Olson & Fazio, 2004), that is, societal views that do not necessarily correspond to the personal view (see later discussion).

CRITICISM OF VALIDATION APPROACHES AT THE INDIVIDUAL LEVEL

As with the validation approach at the group level, the conclusiveness of the validation approach at the individual level is also limited (see Borsboom, Mellenbergh, & van Heerden, 2004). Correlations between IAT effects (e.g., an aggressiveness IAT effect) and criterion variables (e.g., aggressive behavior) might emerge because of a third variable (e.g., impaired cognitive skills) that influences both the IAT effect and the criterion variable. For example, the finding that the IAT predicts aggressive behavior does not necessarily attest to the IAT's validity as a measure of implicit aggressiveness. This correlation might also result from individual differences in cognitive skills such as the ability to inhibit impulsive responses that may impact on both aggressive behavior and the IAT effect (see the IAT's confound by cognitive abilities as discussed later). All in all, however, the evidence provided by correlational studies is, to a large extent, in line with the assumption that IAT effects can capture meaningful construct-related variance.

CRITICISM OF THE IAT

In summary, the IAT has been shown to capture valid construct-related variance with regard to both the group and individual levels. Although these findings are encouraging and indicate the IAT's validity, several studies have seriously challenged the assumption that IAT effects are driven primarily by the to-be-measured associations (for general criticism of the IAT, see Fiedler, Messner, & Bluemke, 2006). First and foremost, researchers criticized that, in contrast to the explosion of IAT applications in diverse psychological areas, the processes underlying the IAT are still unclear (e.g., Fazio & Olson, 2003). For most if not all implicit measures (including the IAT), it is not yet sufficiently understood how the to-be-measured construct translates into observed responses (De Houwer et al., 2009). Identifying the underlying processes of the IAT is particularly important because several factors have been found to contribute to the IAT effect independent of the to-be-measured construct and thus cause additional, but construct-unrelated, variance in the IAT effect (e.g., Sherman et al., 2008; Wentura & Rothermund, 2007). Hence, although the construct in question may be sufficient to cause an IAT effect, other factors might also lead to IAT effects independently of this construct, thereby affecting not only the absolute size but possibly also the rank order of IAT effects.

Unfortunately, there is no comprehensive, testable process model that takes all confounding factors into account and allows their relative influences on the IAT effect to be disentangled. Not even modeling approaches (see Scoring Procedures section) permit the integration of all such factors. However, process models have been proposed that account for at least some, albeit not all, factors that can cause systematic variance in IAT effects. A brief overview of these models is given next.

Process Models of the IAT

Stimulus–Response Compatibilities

According to De Houwer (2001, 2003b), the IAT effect is based on stimulus–response compatibility. The basic assumption in this model is that response keys acquire the meaning of the stimulus category to which they are assigned. Compatibility between the meaning of a response key and stimulus features then facilitates responses. This mechanism can explain the IAT effect because compatibility between stimulus features and responses is consistently given in the compatible, but not the incompatible, block.

To illustrate this process, consider the racial attitude IAT introduced previously. By asking participants to press one key for negative words and another key for positive words, the a priori neutral keys are assumed to become associated with negative and positive valence, respectively (see also Eder & Rothermund, 2008). Hence, for prejudiced individuals who like white persons but dislike black persons, stimuli and responses are compatible (i.e., associated with the same valence) when the "negative" key has to be pressed for black names and the "positive" key for white names (black/negative–white/positive block). When the same individuals are asked to press the "negative" key for white names and the "positive" key for black names (white/negative–black/positive block), stimuli and responses are incompatible. Because stimulus–response compatibility varies between the compatible and incompatible blocks of an IAT, De Houwer (2001, 2003b, 2008) hypothesized that IAT effects are due to the activation of responses by the presented stimuli.

It is important to note that not only relevant stimulus features (such as category membership) but also irrelevant stimulus features (such as perceptual form) might activate responses in an IAT (De Houwer, 2008; De Houwer, Geldof, & De Bruycker, 2005). Thus, the feature according to which stimulus and response in an IAT are compatible might not necessarily (and exclusively) be construct related and thus relevant (such as valence in the prior example), but could also be construct unrelated and thus irrelevant (i.e., features other than valence in attitude IATs such as, e.g., perceptual form; De Houwer et al., 2005). The stimulus–response compatibility account thereby predicts that both construct-related and construct-unrelated variance can contribute to the IAT effect.

Random-Walk Model

Brendl and colleagues (2001) proposed that the IAT effect reflects the result of a random-walk process in which evidence is accumulated on a joint response-related decision dimension. The time required before a response criterion is reached depends on whether all incoming information pushes an internal counter in the same direction. It is hypothesized that both information of the target categories (i.e., category membership such as black vs. white) and information of the attribute categories (e.g., valence) drive the counter. Therefore, stimuli of the target categories (e.g., black vs. white names) should have a lower net accumulation rate in the incompatible than in the compatible IAT condition because information of the category membership (i.e., black vs. white) and valence of a stimulus (i.e., negative vs. positive) disagree in the former, but not in the latter, condition.

Again, let us take the example of the racial attitude IAT. For individuals with implicit prejudices against blacks, black stimuli do not only belong to the category black but are also negatively evaluated. If a black stimulus has to be categorized in the compatible block (here, black/negative vs. white/positive), both sources of information (i.e., the membership of the category black as well as the negative valence) push the accumulation process toward the same response (i.e., the common response for black names and negative words). In contrast, in the incompatible block (here, white/negative vs. black/positive), the two sources of information move the accumulation process in opposite directions because now black names and negative words are to be mapped onto different responses. Usually category membership will have the stronger impact resulting in correct responses in most trials. All in all, however, the net evidence accumulation rate for black stimuli should be lower in the incompatible block than in the compatible block, thus leading to slower responses in the former than in the latter task.

Brendl and colleagues (2001) predicted that differences in net accumulation rate are accompanied by a shift in response criteria in the incompatible block of an IAT. The authors assume that because the incompatible block is perceived as more difficult, participants adopt a more conservative response criterion, leading to slower responses in the incompatible block compared with the compatible block. Accordingly, Brendl and colleagues suggest two mechanisms by which IAT effects are produced, namely, different rates of information accumulation and different response criteria. Whereas Brendl and colleagues did not mathematically formalize their conceptualization of a random-walk model for the IAT, Klauer and colleagues (2007) applied diffusion model analyses to IAT data and found evidence for the two proposed mechanisms. They also confirmed Brendl and colleagues' assumption that the first mechanism, different rates of information accumulation, produces construct-related variance in the IAT effect, whereas the second mechanism, different response criteria, produces construct-unrelated variance of strategic, situational, and/or trait-related influences on the IAT effect.

Task Switching

According to Mierke and Klauer (2001, 2003; Klauer & Mierke, 2005), task-switching costs contribute to the IAT effect because they affect the two crucial blocks of the IAT asymmetrically. Thus, the central assumption of the task-switching account is that the IAT involves executive control processes (i.e., identifying and switching to the appropriate task set). It is argued that in the compatible block of an IAT the structure of the task provides participants with an overlapping feature. Again, think of prejudiced individuals who like white persons but dislike black persons. For these individuals, negative words and black names share the feature negativity, whereas positivity is shared by positive words and white names. In the black/negative–white/positive block (here the compatible block) of the racial attitude IAT, categories that share a feature, namely valence, are thus mapped onto one response key. Categorizing a black or white stimulus according to valence (negative or positive) or category membership (black or white) should thus lead to the same response (Mierke & Klauer, 2001, 2003). Consequently, the task-switching account assumes that participants derive their responses from an arbitrary feature (not necessarily valence) shared by the attribute and target category in the compatible block (see also De Houwer et al., 2005). Because the process of deriving responses is thereby simplified, responses should be fast in this condition.

In contrast, responses cannot be derived from an overlapping feature in the incompatible IAT block. For instance, if the same prejudiced individuals complete the white/negative–black/positive block of an IAT, responding to a black name on the basis of its valence (here negative) would lead to an incorrect response. In the incompatible block, attribute-related information thus needs to be ignored for stimuli of the target categories but has to be processed for stimuli of the attribute categories. Hence, in the incompatible block, participants are required to perform every switch between attribute and target discrimination tasks, whereas the compatible block can be completed without performing all such task switches. Because task switching is associated with performance costs (Rogers & Monsell, 1995) and affects both blocks asymmetrically, task-switching ability should contribute to the IAT effect (see Klauer, Schmitz, Teige-Mocigemba, & Voss, in press, for evidence).

The task-switching account thereby predicts that both construct-related and construct-unrelated variance contribute to the IAT effect: An attitude IAT effect, for instance, should comprise construct-related variance inasmuch as participants derive their responses from valence as the arbitrary feature shared by the attribute and target category in the compatible block. Construct-unrelated variance should constitute the IAT effect inasmuch as participants derive their responses from an arbitrary feature other than valence. Also, construct-unrelated influences of task-switching ability should be larger the more participants tend to save costly task switches in the compatible block by simplifying the task via deriving responses from some overlapping feature.

Figure–Ground Asymmetry

According to Rothermund and Wentura (2001, 2004), the IAT measures differences in the salience of stimulus categories. Figure–ground asymmetries within the target (e.g., black vs. white) and attribute (e.g., negative vs. positive) dimensions are the central explanatory concept of this account. The authors assume that participants simplify, either spontaneously or strategically (Rothermund, Wentura, & De Houwer, 2005), the compatible block, in which the salient categories are mapped onto one response key, by recoding both categorization tasks as figure–ground discriminations. This way, all salient (i.e., figure) stimuli are assigned to one key and all nonsalient (i.e., ground) stimuli to the other, so that the salient stimuli constitute the figure against the background of the less salient stimuli. Importantly, such a recoding is impossible in the incompatible block in which the salient categories are mapped onto different response keys. Hence, performance differences between the two blocks are argued to be the result of salience asymmetries.

Applying the figure–ground asymmetry account to the racial attitude IAT, it may be assumed that black names are more salient than white names because they are unfamiliar. Simultaneously, negative words are more salient than positive words because of the attention-grabbing power of negative information (Pratto & John, 1991). Hence, participants should respond faster and more accurately if the salient categories black and negative share one response (black/negative–white/positive block) than if the salient categories black and negative are mapped onto different responses (white/negative–black/positive block). This is because in the former, but not in the latter, case participants can reduce the complex four-to-two categorization task to a single binary decision of whether the stimulus belongs to the salient (i.e., figure) or the nonsalient (i.e., ground) category. As in Mierke and Klauer's (2001, 2003) task-switching account, the figure–ground asymmetry account thus assumes that the compatible, but not the incompatible, block of an IAT provides participants with an overlapping feature shared by the target and attribute category that directly contributes to the IAT effect (see also De Houwer et al., 2005). In contrast to the task-switching model, however, it is assumed that the feature used for recoding the IAT task must be salience. Hence, associations between categories are argued to play a subordinate role in the IAT effect compared with salience asymmetries (but see Kinoshita & Peek-O'Leary, 2006).

Note, however, that the figure–ground asymmetry account still allows for an influence of valence associations inasmuch as they may serve as determinants of salience asymmetries. For instance, for a prejudiced individual, stimuli of the category black such as names typical for blacks might be salient not only because they are unfamiliar but also because they are negatively valenced (remember that negative stimuli are assumed to be salient). This should lead to better performance when black and negative (i.e., figure) stimuli share one response. Thus, inasmuch as valence is one determinant of salience, salience may serve as a mediator between valence and IAT scores. Consequently, according to the figure–ground asymmetry account, IAT effects might comprise both construct-related variance (if valence of the target categories is a determinant of salience) and construct-unrelated variance (if factors other than valence determine salience).

Summary

Despite the absence of a comprehensive process model, fruitful proposals have been made about the processes by which variables may cause variations in IAT effects. As discussed later, research has confirmed some, albeit not all, predictions of the respective process models. We do not evaluate strengths and weaknesses of every account here. Instead, we focus on a shared assumption underlying all process models, namely that IAT effects are influenced not only by the to-be-measured associations between categories but also by other construct-unrelated factors.

Confounding Factors of the IAT Effect

Unfortunately, in the absence of a comprehensive process model, the relative contribution of construct-related and construct-unrelated influences on the IAT effect cannot be determined and thus cannot be controlled for statistically. Confounding factors, therefore, exert an uncontrollable influence on the absolute size and possibly also on the rank order of IAT effects. Accordingly, the interpretation of IAT effects in an absolute manner is compromised: An IAT effect of zero cannot be interpreted as reflecting a neutral attitude, nor does a positive (negative) IAT effect necessarily reflect a positive (negative) attitude. Thus, descriptive statements such as "72% of test takers show anti-black bias" cannot be derived from IAT effects.

Given that some confounding factors (e.g., cognitive abilities) differ between individuals, they do not exert the same influence on IAT effects for all individuals but rather exert interindividually different influences. Accordingly, such interindividually differing contaminants should distort not only the size but also the rank order of IAT effects, thereby restricting the IAT's predictive power and, hence, affect conclusions about validity. In the absence of empirical evidence, it is difficult to predict whether or not a specific contaminant will affect the rank order of IAT effects. In principle, all of the confounding factors discussed next might distort the rank order, at least to the extent to which participants systematically differ in being subject to the respective contaminant. Confounding factors thus pose a problem for both the interpretation of absolute IAT effects and the interpretation of the IAT effect as a measure of interindividual differences.

In the following, we discuss the different factors that have been shown to contaminate the IAT effect. If available, we also present possible remedies for the respective confounds, some of which involve procedural changes to the IAT. Remarkably, research on IAT contaminants has promoted the development of several implicit measures (for an overview, see Table 7.2) that are conceptually similar to the IAT but aim for overcoming one or more of its confounds. We start the section with a more general (structural) problem of the IAT, before we concentrate on specific confounding factors.

Recoding

Many of the confounding influences on the IAT effect may be argued to have their roots in so-called *recoding processes* (see Rothermund, Teige-Mocigemba, Gast, & Wentura, 2009). As just elaborated, recoding in the IAT means that participants simplify—spontaneously or strategically—one of the double-discrimination tasks of an IAT. Instead of following the instructions to categorize all stimuli according to their category membership, participants may recode the four categories of an IAT into two. Such a simplification may rely on any feature that helps to distinguish between the two groups of stimuli that are assigned to different response keys (Mierke & Klauer, 2003). Participants might even draw on societal views (i.e., extrapersonal knowledge; Olson & Fazio, 2004) to simplify the task (Rothermund & Wentura, 2004). Hence, the IAT effect might, in part, reflect those overlapping features that participants (decided to) use for categorization (e.g., valence, meaning, salience, perceptual form; De Houwer et al., 2005).

Recently, it has been argued that recoding processes rely heavily on the IAT's block structure (Rothermund et al., 2009; Teige-Mocigemba, Klauer, & Rothermund, 2008; see also De Houwer, 2003a). It is assumed that the different mappings of categories onto response keys that are implemented in the compatible versus incompatible block of the IAT promote different processes in the two blocks. Because the IAT effect is based on a comparison of performance in the two separate IAT blocks, such processing differences have a direct impact on the IAT effect. A straightforward remedy thus seems to be the elimination of the IAT's block structure, as has been realized in two paradigms called Single Block IAT (SB-IAT; Teige-Mocigemba et al., 2008) and Recoding Free IAT (IAT-RF; Rothermund et al., 2009). The basic principle of both the SB-IAT and the IAT-RF is that the mapping of categories onto response keys may randomly change from trial to trial instead of blockwise.

THE SB-IAT

The SB-IAT provides participants with a structural feature, namely word position, which signals the mapping of categories onto response keys (i.e., compatible vs. incompatible) for each trial. All stimuli are randomly presented above or below a dashed line that divides the screen into an upper and a lower half. If, for instance, a stimulus appears in the upper half, the compatible mapping is valid (i.e., compatible categories share one response). If a stimulus appears in the lower half, the incompatible mapping is valid (i.e., incompatible categories share one response).

To illustrate the procedure, let us consider a racial attitude SB-IAT: For attribute stimuli (e.g., positive and negative words), word position is irrelevant because attribute stimuli always have to be assigned to the same response keys irrespective of word position (e.g., positive words to the right key, negative words to the left key). For target stimuli (e.g., names typical for whites vs. blacks), however, word position is highly relevant. For instance, if target stimuli appear above the dashed line, names of whites have to be assigned to the right (positive) key, whereas names of blacks have to be assigned to the left (negative) key. Conversely, if target stimuli appear below the dashed line, names of whites have to be assigned to the

TABLE 7.2. Overview of Measures Reviewed in the Present Chapter, Including Their Procedural Differences to the IAT, the IAT Confound They Address, and Their Possible Caveat

Measure	Procedural differences from the IAT	Problem addressed	Caveat
SB-IAT	Eliminates IAT's block structure	Recoding, method-specific variance, compatibility order	Sparse empirical evidence at this point
IAT-RF	Eliminates IAT's block structure	Recoding, task-switching costs, stimulus influences, compatibility order	Sparse empirical evidence at this point
IAP	Requires approach–avoidance behavior by changing the response modality (joystick movement instead of key press)	Strategic control (?)	Sparse empirical evidence at this point, implementation effort higher
EMA	Requires approach–avoidance behavior by using horizontal joystick movement instead of key press	Relative measure	See IAP; does not allow for between-individual comparisons
Personalized IAT	Removes error feedback, uses more personalized attribute category labels, requires participants to categorize attribute stimuli according to (explicit) personal preference (instead of normative valence)	Extrapersonal associations	Vague conceptualization of extrapersonal associations; more direct measure
EAST	Differs in many aspects, main difference: Uses "valenced" responses for color discrimination task of target words, EAST effect relies on stimulus–response compatibility of task-irrelevant valence of target words and valence of responses	Recoding, relative measure, compatibility order, limitation to assessment of single constructs	Insufficient psychometric properties in most domains
ID-EAST	Similar to EAST, main difference: requires processing of target categories	See EAST	Sparse empirical evidence at this point
SC-IAT	Uses three categories instead of four (one target, two attribute categories)	Relative measure	Recoding (?)
GNAT	Uses go/no-go task, GNAT scores typically based on signal detection analyses of error rates	Relative measure	Lower reliability
SPF	Eliminates IAT's block structure, requires categorizing two stimuli (one target, one attribute) at once to one of four possible target–attribute category pairs	Relative measure, compatibility order, recoding (?)	Sparse empirical evidence at this point
BIAT	Uses substantially fewer trials, requires focusing on just two categories (one target, one attribute category) in each four-category test block	Time required for task	Sparse empirical evidence at this point

Note. IAT, Implicit Association Test; SB-IAT, Single-Block IAT; IAT-RF, Recoding-Free IAT; IAP, Implicit Association Procedure; EMA, Evaluative Movement Assessment; EAST, Extrinsic Affective Simon Task; ID-EAST, Identification-EAST; SC-IAT, Single Category IAT; GNAT, Go/No-Go Association Test; SPF,= Sorting Paired Features Task; BIAT, brief IAT.

	…'s block structure, requires categorizing two stimuli (one target, one attribute) at once to one of four possible target–attribute category pairs	Relative measure, compatibility order, recoding (?)	Sparse empirical evidence at this point
BIAT	Uses substantially fewer trials, requires focusing on just two categories (one target, one attribute category) in each four-category test block	Time required for task	Sparse empirical evidence at this point

Note. IAT, Implicit Association Test; SB-IAT, Single-Block IAT; IAT-RF, Recoding-Free IAT; IAP, Implicit Association Procedure; EMA, Evaluative Movement Assessment; EAST, Extrinsic Affective Simon Task; ID-EAST, Identification-EAST; SC-IAT, Single Category IAT; GNAT, Go/No-Go Association Test; SPF,= Sorting Paired Features Task; BIAT, brief IAT.

left (negative) key and names of blacks have to be assigned to the right (positive) key. As in the IAT, the performance difference between the two kinds of mappings is interpreted as reflecting the relative association strength between the target and attribute categories. For individuals with implicit prejudices against blacks, responses to stimuli that appear in the upper half of the screen (white/positive and black/negative share one response) should thus be faster and more accurate than responses to stimuli appearing in the lower half (black/positive and white/negative share one response).

THE IAT-RF

In the IAT-RF, a structural feature such as word position that signals the mapping of categories onto response keys for each trial is absent. Instead, response assignments are indicated at the beginning of each trial by presenting the category labels in the respective left and right corners of the screen (e.g., black/negative left, white/positive right) that correspond to the response keys to which the categories are assigned in the upcoming trial. For each trial, participants are thus required to read the category labels that indicate the mapping for the respective trial.

The main difference between the SB-IAT and the IAT-RF on the one hand and the standard IAT on the other is that both the SB-IAT and the IAT-RF compare performance on compatible versus incompatible trials within the same (i.e., a single) block, whereas the standard IAT compares performance on compatible versus incompatible trials between two different (i.e., compatible vs. incompatible) blocks. Thus, in the SB-IAT and the IAT-RF, the response mapping (compatible vs. incompatible) may randomly change from trial to trial and is not consistently blocked anymore. This should impede any kind of recoding strategies because recoding processes are assumed to rely on a consistent mapping of categories onto response keys (Strayer & Kramer, 1994).

The SB-IAT and the IAT-RF indeed showed reduced susceptibility to markers of recoding processes (see later discussion): Confounding influences on the IAT such as those of method-specific variance, task-switching costs, and biased selection of stimuli were clearly diminished (Rothermund et al., 2009; Teige-Mocigemba et al., 2008), suggesting usefulness and effectiveness of eliminating the IAT's block structure. At the same time, first findings indicated satisfactory reliability and validity estimates at both the group (e.g., known group approach) and the individual (implicit–explicit consistency) level. It should be noted, however, that effect sizes of both the SB-IAT and the IAT-RF were clearly reduced compared with the IAT, thereby leaving less room for markers of recoding processes to appear. Definitely, more research is needed to evaluate the potential of both the SB-IAT and the IAT-RF.

Cognitive Abilities

Maybe best documented and acknowledged is the confounding influence of cognitive abilities on the IAT effect. For instance, overall response speed and the size of IAT effects have been found to be correlated (McFarland & Crouch, 2002). Because overall response speed is associated with cognitive abilities, these results suggest that IAT effects are at least partially determined by the participants' cognitive skills. Indirect evidence comes from studies showing larger IAT effects for older individuals compared with younger individuals (e.g., Hummert, Garstka, O'Brien, Greenwald, & Mellott, 2002). Given that cognitive abilities tend to decline with age, such findings also suggest that IAT effects are influenced by cognitive abilities (see also Gonsalkorale, Sherman, & Klauer, 2009; Sherman et al., 2008). Further indirect evidence for a cognitive skill confound on the IAT is provided by studies showing correlations between different IATs that were supposed to capture different, unrelated constructs and, therefore, should not be intercorrelated (e.g., Back, Schmukle, & Egloff, 2005; McFarland & Crouch, 2002; Mierke & Klauer, 2003). Such findings indicate that some of the systematic variance in IAT effects is due to factors that affect different IATs similarly, irrespective of contents, reflecting so-called method variance. Method-specific variance in the IAT has been accounted for by speed–accuracy trade-offs (Klauer et al., 2007) and by cognitive abilities, in particular task-switching ability (Klauer et al., in press), as is to be expected following Mierke and Klauer's (2003) task-switching model.

Different techniques have been proposed to decrease the cognitive skill confound on the IAT effect. First, *D* measures (Greenwald et al., 2003) have proven to be less susceptible to cognitive skills than the conventional score. Second, modeling approaches (e.g., the Quad model; Conrey et al., 2005) may help to dissociate construct-related and construct-unrelated components of the IAT effect (see also Klauer et al., 2007), including, for instance, a component for cognitive abilities (Sherman et al., 2008). Finally, the SB-IAT effect and the IAT-RF effect have been found to be less contaminated by method variance (see prior discussion).

Salience

In support of their figure–ground asymmetry account, Rothermund and Wentura (2001, 2004) reported experimental data (i.e., manipulations of salience influence IAT effects) and correlational data (i.e., IAT effects are related to measures of salience). These findings corroborate the assumption that salience asymmetries have the potential to contribute to IAT effects as acknowledged by the developers of the IAT (see Greenwald, Nosek, Banaji, & Klauer, 2005). It is still controversial, however, how pervasive the impact of salience asymmetries is (see Rothermund et al., 2005). Recent studies indicate that only part of the IAT effect can be accounted for by construct-unrelated salience asymmetries because construct-related compatibilities between the nominal categories have been shown to simultaneously contribute to IAT effects (e.g., Kinoshita & Peek-O'Leary, 2006). Moreover, there is still uncertainty at the conceptual level about how salience should be measured (e.g., Greenwald et al., 2005) and how it is related to other constructs such as familiarity and polarity (e.g., Kinoshita & Peek-O'Leary, 2006; Proctor & Cho, 2006).

Given that recoding processes have been argued to form the basis of contaminations by salience asymmetries (e.g., Wentura & Rothermund, 2007), prevention of recoding processes should also prevent confounding influences of salience on the IAT effect (see Rothermund et al., 2009). Also, measuring salience asymmetries using a visual search task might enable one to estimate its impact on the IAT effect (Rothermund & Wentura, 2004).

Stimuli

The IAT effect has been found to be determined both by the superordinate nominal categories according to which the stimuli have to be categorized (i.e., the category labels such as black vs. white) (De Houwer, 2001, 2008; Olson & Fazio, 2003) and by the stimuli used to represent the categories (e.g., a particular black or white face) (Bluemke & Friese, 2006; Govan & Williams, 2004; Mitchell, Nosek, & Banaji, 2003; Steffens & Plewe, 2001). Influences at the level of the category labels are desired. They ensure the experimenter's control over the nominal categories according to which participants categorize and process the stimuli (see the relevant feature account by De Houwer, 2008). This allows for determining the construct that the IAT effect should reflect, and also adds to the IAT's easy applicability to various domains.

Influences at the level of the stimuli, however, are often unintended. For example, several studies indicated that stimulus selection may force participants to categorize stimuli according to other than the specified category labels (see the irrelevant feature account by De Houwer, 2008). As Govan and Williams (2004) proposed, participants may redefine the category labels in order to reconcile meaning and/or valence of category labels with meaning and/or valence of stimuli. Biased selections of stimuli can thus dramatically influence magnitude and even direction of IAT effects (Bluemke & Friese, 2006; Govan & Williams, 2004; Rothermund et al., 2009), which poses a threat to the IAT's validity.

Careful stimulus selection is thus required to exert as much control as possible over the nominal categories according to which participants categorize stimuli. First and foremost, stimuli should be representative for the respective category, and any confounds of stimulus features of the attribute versus target categories should be avoided (Steffens & Plewe, 2001). In an attitude IAT, for instance, both positively and negatively valenced stimuli should be selected for the target categories (De Houwer, 2001). Furthermore, the distinctiveness of the attribute versus target stimuli may be enhanced by distinct colors, fonts, or other stimulus modalities (Nosek et al., 2007). Finally, recent findings indicated that elimination of the IAT's block structure may be an effective means to reduce stimulus influences (Rothermund et al., 2009), and an IAT variant that uses the category labels (or synonyms of them) as stimuli also provided promising results (Steffens et al., 2008).

Strategic Effects

Evidence for strategic effects on the IAT comes from studies that investigated the fakeability of the IAT. These studies revealed that the IAT outcome can be strategically controlled (1) if participants are told how to fake (Fiedler & Bluemke, 2005), (2) if participants are high on self-monitoring and highly motivated to fake (Czellar, 2006), or (3) if participants had experience with at least one prior IAT (Fiedler & Bluemke, 2005; Steffens, 2004). If, however, participants were exposed to an IAT for the very first time (Banse et al., 2001; but see

De Houwer, Beckers, & Moors, 2007) or were not advised on how to fake (Asendorpf et al., 2002; Egloff & Schmukle, 2002; but see Experiment 3 of Lowery, Hardin, & Sinclair, 2001), there was little evidence for strategic control over the IAT outcome. Accordingly, under certain circumstances, participants might strategically influence the IAT effect. It is, however, probably much easier to exert strategic control over self-reports than over an IAT (Steffens, 2004).

To the best of our knowledge, no straightforward solution to control for faking attempts in IATs has been proposed so far. Promising approaches may comprise the development of algorithms that allow for distinguishing fakers from nonfakers or the use of modeling approaches as discussed previously. Furthermore, procedural changes might reduce the risk of strategically altered IAT effects. For instance, modifying the Evaluative Movement Assessment (EMA; Brendl, Markman, & Messner, 2005),[2] Schnabel, Banse, and Asendorpf (2006b) used what they called the Implicit Association Procedure (IAP).

THE IAP

The IAP is methodologically very similar to the IAT, the main difference being the use of a different response modality (i.e., joystick movement instead of key press) that is thought to trigger approach (pulling the joystick toward oneself) and avoidance (pushing the joystick away from oneself) behavior. In an initial combined block of a racial attitude IAP, for instance, participants have to pull the joystick to themselves when white or positive stimuli are presented and to push it away from themselves when black or negative stimuli are presented. In the reverse combined block, participants have to pull the joystick to themselves when black or positive stimuli are presented and to push it away from themselves when white or negative stimuli are presented.

As in the IAT, the difference in performance between the two kinds of pairings is interpreted as revealing the relative association strength between the target and attribute categories. Accordingly, individuals with implicit prejudices against blacks are expected to respond faster and more accurately when black and negative stimuli require avoidance behavior (pushing the joystick away from oneself) and white and positive stimuli require approach behavior (pulling the joystick toward oneself) compared with the reverse configuration. The IAP showed satisfactory psychometric properties, and unlike the IAT, it was not susceptible to faking (Schnabel et al., 2006b). This might indicate that it is more difficult to exert strategic control over the outcome of a procedure that uses approach–avoidance responses.

Extrapersonal Associations

Olson and Fazio (2004) identified another confounding influence on the IAT effect: so-called "extrapersonal" associations (see also Karpinski & Hilton, 2001). The term *extrapersonal knowledge* refers to culturally shared assumptions (e.g., apples are healthy and thus are positive) that do not necessarily correspond to personal evaluations (e.g., I don't like apples). To the extent that the measurement purpose is to reveal personal rather than societal views, the IAT's sensitivity to extrapersonal associations poses a threat to its validity (Olson & Fazio, 2004; but see Nosek & Hansen, 2008a).

Evidence for the IAT's contamination by extrapersonal associations is provided by experiments in which the manipulation of extrapersonal views changed IAT effects (Han, Olson, & Fazio, 2006). Furthermore, when groups with diverging personal and societal views completed IATs, IAT effects at least sometimes seemed to be in line with societal views (e.g., Olson & Fazio, 2004; Spruyt et al., 2007). Finally, the assumption that extrapersonal associations contaminate IAT effects is consistent with studies showing the IAT's weakness to predict behavior in health-related domains (e.g., Spruyt et al., 2007), where societal views are prevalent. To prevent a confounding influence of extrapersonal associations, Olson and Fazio (2003) suggested procedural changes to the IAT. In their personalized IAT variant, error feedback is removed, more personalized attribute category labels are used (i.e., "I like" vs. "I don't like" instead of "positive" vs. "negative"), and participants are asked to categorize attribute stimuli according to their (explicit) personal preference (instead of normative valence). Although the personalized IAT was less affected by societal views (Han et al., 2006; Olson & Fazio, 2004), recent research seriously questioned the usefulness of this IAT variant as an implicit measure: Nosek and Hansen (2008b) showed that personalized IATs foster participants to evaluate not only attribute stimuli but also target stimuli explicitly, thereby basically making the measurement more direct on an indirect–direct dimension.

Moreover, doubts have been raised about the validity and theoretical significance of the extrapersonal account of IAT effects. At the empirical level, recent correlational studies provided little evidence for a link between IAT effects and measures of societal views (Nosek & Hansen, 2008a). At the conceptual level, it has been argued that the distinction between personal and extrapersonal views actually makes little sense, especially when considering the automatic effects of personal and extrapersonal associations (Gawronski & Bodenhausen, 2006; Nosek & Hansen, 2008a). Furthermore, there is still uncertainty about how extrapersonal associations can be conceptualized (see Gawronski, Peters, & LeBel, 2008). Finally, the processes via which extrapersonal associations may contaminate IAT effects are not yet identified (but see Rothermund & Wentura, 2004, for a strategic recoding account).

Compatibility Order

The IAT effect is also known to be confounded by compatibility order: IAT effects tend to be larger if the compatible block precedes the incompatible block than vice versa (see Nosek et al., 2007). A theoretical account for compatibility-order effects was provided by Klauer and Mierke (2005). Drawing on their task-switching model, the authors suggested that differences in the accessibility of attribute information in the compatible versus incompatible block of the IAT may account for compatibility-order effects. Such effects are difficult to control for, given that compatibility is a function of interindividual differences in the construct of interest and cannot a priori be determined in many applied contexts. First attempts to reduce the confounding impact of compatibility order have focused on slight changes to the IAT procedure (Nosek, Greenwald, & Banaji, 2005), namely by increasing numbers of trials in the reversed target discrimination task (fifth block of Table 7.1). Also, implicit measures that abandon the IAT's block structure such as the SB-IAT, the IAT-RF, or the Extrinsic Affective Simon Task (EAST; De Houwer, 2003a) should not be subject to compatibility-order effects (Teige-Mocigemba et al., 2008).

THE EAST

In the EAST, participants are asked to respond to attribute words that are colored white and to target words colored blue or green by pressing one of two keys. For the white attribute words (e.g., *positive, negative*), participants have to respond based on valence. As a result, one key is assumed to become associated with positive valence (positive key) and the other key with negative valence (negative key). For the colored target words (e.g., *blacks*), participants need to select one of those same valenced responses but this time based on word color while ignoring valence of the word. Results typically show that participants perform better when the irrelevant valence of a colored target word corresponds to the valence of the response than when stimulus and response have a different valence (De Houwer, 2003a).

To illustrate the procedure, let us consider a racial attitude EAST: In such an EAST application, attribute words might comprise positive and negative words (e.g., *happy, sad*) that are presented in white color, and target words might comprise names typical for blacks versus whites that are presented in blue color on some trials and in green color on others, respectively. Participants are to press a left key when they see a white word of negative valence (e.g., *sad*) or a word printed in blue (i.e., black or white name) and to press a right key when they see a white word of positive valence (e.g., *happy*) or a word printed in green (i.e., black or white name). To the degree that participants show faster or more accurate responses to a colored (target) word (e.g., black name) when the required response is combined with a negative compared with a positive response, it is inferred that participants have negative associations with the object depicted by the colored word. Accordingly, individuals with implicit prejudices against blacks are expected to respond faster and more accurately when the word color of black stimuli requires them to press the negative key compared with the positive key.

Although in some domains the EAST provided promising results (e.g., De Houwer & De Bruycker, 2007c), several studies questioned its usefulness as a measure of interindividual differences and showed clearly inferior psychometric properties compared with the IAT (e.g., De Houwer & De Bruycker, 2007b; Teige et al., 2004). In part, the EAST's unsatisfactory reliability could be improved by some procedural variations: The so-called identification-EAST (ID-EAST; De Houwer & De Bruycker, 2007a), for instance, differs from the EAST in that it requires participants to process the target categories in order to perform the task. The few studies that used the ID-EAST so far (De Houwer & De Bruycker, 2007a; Rudolph et al., 2008) indeed found that the psychometric properties for this variant were superior to the EAST, although still inferior to the IAT. Given the sparse empirical data on the ID-EAST, however, it appears to be premature to evaluate the usefulness of this EAST variant at present.

Relative Measure

Last but not least, it has been criticized that the IAT is restricted to the assessment of relative association strengths between nominal categories. For instance, a positive score in a racial attitude IAT does not indicate that one evaluates blacks negatively and whites positively. It rather reflects that blacks are evaluated more negatively than whites. Thus, an IAT effect does not permit conclusions about a person's evaluation of the single categories, nor does the same IAT effect of different persons necessarily reflect the same attitude (e.g., Blanton, Jaccard, Gonzales, & Christie, 2006). As a second problem, the IAT's applicability is limited to constructs that have a natural counterpart. Several constructs of interest, however, do not meet this requirement. For example, if researchers are interested in a person's fear of spiders (e.g., Teachman, 2007), it is difficult to think of a suitable counterpart that could serve as a contrast category. Research has, therefore, suggested different solutions such as contrasting the target category with a neutral category (e.g., Sherman et al., 2003) or using other implicit measures that allow for the assessment of associations between a single-target category and attribute categories. Among these measures are the Single Category IAT (SC-IAT; Karpinski & Steinman, 2006), the Go/No-Go Association Test (GNAT; Nosek & Banaji, 2001), the Sorting Paired Features Task (SPF; Bar-Anan, Nosek, & Vianello, 2009), and the EAST (De Houwer, 2003a).

THE SC-IAT

The SC-IAT is very similar to the IAT, except that it uses three categories instead of four (one target, two attribute categories): The combined blocks involve assigning two categories to one response and one category to the other response (e.g., black/negative left, positive right or negative left, black/positive right). Although reliability of the SC-IAT has been shown to be somewhat lower than that of the IAT (see Schnabel et al., 2008), validity estimates are encouraging (Bluemke & Friese, 2008; Friese, Hofmann, & Schmitt, 2008). However, because of its high methodological similarity to the IAT, many of the confounding influences on the IAT may also affect SC-IATs (Karpinski & Steinman, 2006).

THE GNAT

The GNAT uses a go/no-go task in which participants are first asked to show a "go" response (e.g., key press) to stimuli of a target and an attribute category (e.g., blacks and negative in a racial attitude GNAT) and a "no-go" response (i.e., no key press) to distracter stimuli, some of which belong to the opposed attribute category (here positive). In a second block, the responses for the attribute categories are reversed such that the former go category (here negative) becomes the no-go category and the former no-go category (here positive) becomes the go category. Accordingly, in the second block of a racial attitude GNAT, participants are to show a go response to black and positive stimuli and a no-go response to distracter stimuli, including negative stimuli. Go responses typically have to be made within a given response deadline (e.g., 600 msec).

The performance difference between the two pairings of go trials (here black/negative vs. black/positive) is then interpreted as reflecting the association between the target category and the attribute categories. Note that, in contrast to the other implicit measures discussed in the present chapter, GNAT scores are based on signal detection analyses of error rates rather than response latencies (for details, see Nosek & Banaji, 2001). Thus, individuals with implicit prejudices against blacks are expected to perform better in terms of sensitivity scores (d') when black and negative stimuli require a go response compared with when black and positive stimuli require a go response. As with the SC-IAT, the GNAT's reliability has been found to be lower than the IAT's (e.g., Nosek & Banaji, 2001). However, some studies have attested to the GNAT's validity (e.g., Boldero, Rawlings, & Haslam, 2007; Gonsalkorale, von Hippel, Sherman, & Klauer, 2009; Teachman, 2007).

THE SPF

Similar to the SB-IAT and the IAT-RF, the SPF eliminates the IAT's block structure. Unlike the former two tasks, the SPF uses four response options instead of two: Each trial requires participants to categorize two stimuli (one target, one attribute) at once to one of four possible target–attribute category pairs, thereby forcing the simultaneous processing of both (i.e., target and attribute) association components. In a racial attitude SPF, for instance, the category pairs whites–good, whites–bad, blacks–good, and blacks–bad are each mapped onto one response key. If, for example, a name typical for whites (target stimulus) and the word *wonderful* (attribute stimulus) are simultaneously presented, participants are to press the "whites–good" key. If a stimulus pair consists of a name typical for whites and the word *awful*, participants are to press the "whites–bad" key and so on.

According to Bar-Anan and colleagues (2009), the SPF has the unique feature of allowing for separable assessments of the four association strengths involved (e.g., whites–good, whites–bad, blacks–good, blacks–bad). Although the authors present some evidence for differential predictive validity of the four specific scores, they acknowledge the interdependence of the four scores: Each association strength can only be interpreted relative to the other three but not in isolation. Initial studies showed that reliability estimates of the SPF are considerably lower than for the IAT but indicated its validity, both at the group level (known-group approach) and at the individual level (implicit–explicit and implicit–implicit consistency; Bar-Anan et al., 2009). Note that the SPF has been shown to be sensitive to both focal, attended concepts (e.g., race in a racial attitude SPF) and nonfocal, unattended stimulus features (e.g., gender of race stimuli in a racial attitude SPF). Future research will have to clarify the extent to which the sensitivity to nonfocal stimulus features is advantageous or disadvantageous (cf. confounding influence of stimulus effects in the IAT).

THE BRIEF IAT

Finally, Sriram and Greenwald (2009) recently introduced the Brief Implicit Association Test (BIAT). The BIAT differs from the IAT in that it uses substantially fewer trials (about one-third the number of IAT trials) and requires participants to focus on just two, so-called focal, of the four categories in each four-category test block. In the initial combined test block of a racial attitude BIAT, for instance, participants have to press one response key when a stimulus matches either the "whites" or the "good" category and have to press the other response key for "anything else" (here blacks and negative stimuli). In the reverse combined test block, the other target category becomes one of the focal categories: Participants thus have to press one response key when a stimulus matches either the "blacks" or the "good" category and have to press the other response key for "anything else" (here whites and negative stimuli). As in the IAT, the performance difference between the initial and the reverse combined blocks is interpreted as reflecting the association between the target categories and the attribute category.

First findings indicate the crucial role of determining which category is focal: Reliability and validity estimates were satisfactory only when "good" (but not "bad") was a focal category in attitude BIATs and when "self" (but not "others") was a focal category in identity BIATs (Sriram & Greenwald, 2009). Note that, in principle, the BIAT may be adapted to allow for the assessment of associations between a single target category and two attribute categories. In the initial combined block of a racial attitude BIAT, for instance, "blacks" and "positive" could serve as the focal categories, whereas "blacks" and "negative" could be focal in the reverse combined block. Clearly, more research is needed to evaluate potentials and limits of the newly developed BIAT, particularly regarding the impact of determining the focal categories on the BIAT's psychometric properties.

CONCLUSION

Since its publication, the IAT has stimulated an enormous amount of fruitful research, revealing its strengths as well as its shortcomings. In the present chapter, we have reviewed this research on the IAT and related tasks. Although many findings regarding the IAT's psychometric properties appear to be promising, the IAT effect has been shown to be contaminated by several variables operating via different processes. We considered different remedies for the respective contaminants, some of which involve procedural changes to the IAT. The review reveals that, after 11 years of IAT research, many questions are still open, obliging one to show scientific responsibility in using the IAT (and any of its derivates) for individual diagnosis (e.g., for employee selection). Nevertheless, IAT research has clearly contributed to a deeper understanding of human experience and behavior and will certainly continue to do so in the future.

ACKNOWLEDGMENTS

The development of this chapter was supported by Grant No. Kl 614/13-1 from the Deutsche Forschungsgemeinschaft to Karl Christoph Klauer and by Grant No. BCS 0802855 from the U.S. National Science Foundation to Jeffrey W. Sherman.

fference between the ini-
ined blocks is interpreted
tion between the target
ite category.
the crucial role of deter-
focal: Reliability and va-
factory only when "good"
ical category in attitude
but not "others") was a
BIATs (Sriram & Green-
in principle, the BIAT
or the assessment of asso-
target category and two
e initial combined block
, for instance, "blacks"
: as the focal categories,
gative" could be focal in
k. Clearly, more research
ntials and limits of the
rticularly regarding the
focal categories on the
erties.

IAT has stimulated an
il research, revealing its
rtcomings. In the pres-
ed this research on the
ough many findings re-
etric properties appear
effect has been shown
ral variables operating
e considered different
contaminants, some of
ianges to the IAT. The
years of IAT research,
n, obliging one to show
sing the IAT (and any
ial diagnosis (e.g., for
theless, IAT research
deeper understanding
behavior and will cer-
he future.

was supported by Grant
che Forschungsgemein-
and by Grant No. BCS
Science Foundation to

NOTES

1. Note that there is some ambiguity in the use of the terms *implicit* and *explicit*. In the literature, the term *implicit measure* is commonly used to refer to indirect measures such as response-time measures. The term *explicit measure* is commonly used to refer to direct measures such as questionnaires or other kinds of self-reports. It has been argued that this terminology makes it difficult to disentangle the empirical measurement level and the construct level. As a result, the outcome of direct and indirect measures might be inadmissibly equated with the underlying implicit and explicit constructs (De Houwer & Moors, 2007; Fazio & Olson, 2003). Furthermore, it is rarely defined what the term *implicit measure* actually means, which leads to a somewhat arbitrary use of the term (for an attempt to contribute to definitional clarity, see De Houwer et al., 2009). For these and other reasons, De Houwer (2006) regarded the terms *direct measure* and *indirect measure* as more appropriate.

   Although we acknowledge the problems posed by the wording, we decided to stick to the more commonly used terminology of implicit and explicit measures in the present chapter (see Fazio & Olson, 2003). We use these terms, however, without making specific assumptions about their exact relations to automaticity and/or unconsciousness. In line with the literature, we reserve the term *implicit measure* for indirect computer-based measures (such as the IAT), whereas other indirect measures not necessarily relying on highly accurate computer-based methods (such as projective tests) will be denoted as *indirect measures*.
2. In the EMA, participants are presented with distracter words (i.e., positive and negative words) and target words (i.e., words describing the attitude objects of interest, e.g., "flower") that are displayed to the right or to the left of the participants' name. Using a joystick, participants have to horizontally move the stimuli either toward their own name, which is assumed to trigger approach behavior, or away from it, which is assumed to trigger avoidance behavior. There are two critical blocks in the EMA: In one block target words have to be moved toward the participant's name, whereas in the other block, target words have to be moved away from the participant's name. For distracter words, participants always have to move positive words toward their own name and negative words away from it independent of the block condition. Valence of each target word can thus be estimated as the difference between the response times for moving the respective target word away from the participant's name (avoidance) versus toward the participant's name (approach). Thus, it is assumed that for positively evaluated objects (such as "flower"), it is easier to show approach behavior compared with avoidance behavior. The EMA has been shown to measure rank orders of preferences for different objects *within* the same individual: It allows for assigning multiple attitude objects to a single scale, retaining their rank order on a good–bad dimension, and meaningfully centering this scale around a neutral point (Brendl et al., 2005). Importantly, the EMA thereby differs from the other implicit measures discussed in the present chapter that aim at assessing differences *between* individuals in their evaluation of the same object.

REFERENCES

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

Back, M. D., Schmukle, S. C., & Egloff, B. (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 & J. S. Nairne (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.

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

Bar-Anan Y., Nosek, B. A., & Vianello, M. (2009). The Sorting Paired Features Task: A measure of association strengths. *Experimental Psychology, 56,* 329–343.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6,* 242–261.

Blanton, H., Jaccard, J., Gonzales, P. M., & Christie, C. (2006). Decoding the Implicit Association Test: Implications for criterion prediction. *Journal of Experimental Social Psychology, 42,* 192–212.

Bluemke, M., & Friese, M. (2006). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology, 42,* 163–176.

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

Boldero, J. M., Rawlings, D., & Haslam, N. (2007). Convergence between GNAT-assessed implicit and explicit personality. *European Journal of Personality, 21,* 341–358.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111,* 1061–1071.

Brendl, C. M., Markman, A., & 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.

Brendl, C. M., Markman, A. B., & Messner, C. (2005). Indirectly measuring evaluations of several attitude objects in relation to a neutral reference point. *Journal of Experimental Social Psychology, 41,* 346–368.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: The quad-model of implicit task performance. *Journal of Personality and Social Psychology, 89,* 469–487.

Conrey, F. R., & Smith, E. R. (2007). Attitude representation: Attitudes as patterns in a distributed, connectionist representational system. *Social Cognition, 25,* 718–735.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12,* 163–170.

Czellar, S. (2006). Self-presentational effects in the Implicit Association Test. *Journal of Consumer Psychology, 16,* 92–100.

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

De Houwer, J. (2003b). A structural analysis of indirect measures of attitudes. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 219–244). Mahwah, NJ: Erlbaum.

De Houwer, J. (2006). What are implicit measures and why are we using them? In R. W. Wiers & A. W. Stacy (Eds.), *Handbook of implicit cognition and addiction* (pp. 11–28). Thousand Oaks, CA: Sage.

De Houwer, J. (2008). Comparing measures of attitudes at the functional and procedural level: Analysis and implications. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 361–390). New York: Psychology Press.

De Houwer, J., Beckers, T., & Moors, A. (2007). Novel attitudes can be faked on the Implicit Association Test. *Journal of Experimental Social Psychology, 43,* 972–978.

De Houwer, J., & De Bruycker, E. (2007a). The identification-EAST as a valid measure of implicit attitudes toward alcohol-related stimuli. *Journal of Behavior Therapy and Experimental Psychiatry, 38,* 133–143.

De Houwer, J., & De Bruycker, E. (2007b). The Implicit Association Test outperforms the Extrinsic Affective Simon Task as a measure of interindividual differences in attitudes. *British Journal of Social Psychology, 46,* 401–421.

De Houwer, J., & De Bruycker, E. (2007c). Implicit attitudes toward meat and vegetables in vegetarians and non-vegetarians. *International Journal of Psychology, 42,* 158–165.

De Houwer, J., Geldof, T., & De Bruycker, E. (2005). The Implicit Association Test as a general measure of similarity. *Canadian Journal of Experimental Psychology, 59,* 228–239.

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 179–194). New York: Guilford Press.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135,* 347–368.

Eder, A., & Rothermund, K. (2008). When do motor behaviors (mis)match affective stimuli?: An evaluative coding view of approach and avoidance reactions. *Journal of Experimental Psychology: General, 137,* 262–281.

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. *Journal of Personality Assessment, 84,* 82–88.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). San Diego, CA: Academic Press.

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

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Test. *Basis and Applied Social Psychology, 27,* 307–316.

Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the "I", the "A" and the "T": Logical and psychometric critique of the Implicit Association Test (IAT). *European Review of Social Psychology, 17,* 74–147.

Friese, M., Hofmann, W., & Schmitt, M. (2008). When and why do implicit measures predict behavior? Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology, 19,* 285–338.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692–731.

Gawronski, B., Deutsch, R., LeBel, E. P., & Peters, K. R. (2008). Response interference as a mechanism underlying implicit measures: Some traps and gaps in the assessment of mental associations with experimental paradigms. *European Journal of Psychological Assessment, 24,* 218–225.

Gawronski, B., Peters, K. R., & LeBel, E. P. (2008). What makes mental associations personal or extra-personal? Conceptual issues in the methodological debate about implicit attitude measures. *Social and Personality Psychology Compass, 2,* 1002–1023.

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2009). Aging and prejudice: Diminished regulation of automatic race bias among older adults. *Journal of Experimental Social Psychology, 45,* 410–414.

Gonsalkorale, K., von Hippel, W., Sherman, J. W., & Klauer, K. C. (2009). Bias and regulation of bias in intergroup interactions: Implicit attitudes toward Muslims and interaction quality. *Journal of Experimental Social Psychology, 45,* 161–166.

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.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4–27.

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. *Journal of Experimental Psychology: General, 134,* 420–425.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97,* 17–41.

Gschwendner, T., Hofmann, W., & Schmitt, M. (2008). Differential stability: The effects of acute and chronic construct accessibility on the temporal stability of the Implicit Association Test. *Journal of Individual Differences, 29,* 70–79.

Han, H. A., Olson, M. A., & Fazio, R. H. (2006). The influence of experimentally created extrapersonal associations on the Implicit Association Test. *Journal of Experimental Social Psychology, 42,* 259–272.

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.

Hofmann, W., Gschwendner, T., Nosek, B. A., & Schmitt, M. (2005). What moderates implicit-explicit consistency? *European Review of Social Psychology, 16,* 335–390.

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.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81,* 774–778.

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

Kinoshita, S., & Peek-O'Leary, M. (2006). Two bases of the compatibility effect in the Implicit Association Test (IAT). *Quarterly Journal of Experimental Psychology, 59,* 2102–2120.

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 IAT effect. *Personality and Social Psychology Bulletin, 31,* 208–217.

Klauer, K. C., Schmitz, F., Teige-Mocigemba, S., & Voss, A. (in press). Understanding the role of executive control in the Implicit Association Test: Why flexible people have small IAT effects. *Quarterly Journal of Experimental Psychology.*

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the Implicit Association Test: A diffusion-model analysis. *Journal of Personality and Social Psychology, 93,* 353–368.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility—A model and taxonomy. *Psychological Review, 97,* 253–270.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: IV: What we know (so far) about the method. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 59–102). New York: Guilford Press.

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1,* 27–66.

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.

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.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.

Nier, J. A. (2005). How dissociated are implicit and explicit racial attitudes?: A bogus pipeline approach. *Group Processes and Intergroup Relations, 8,* 39–52.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

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., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics, 6,* 101–115.

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. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). New York: Psychology Press.

Nosek, B. A., & Hansen, J. J. (2008a). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition and Emotion, 22*, 553–594.

Nosek, B. A., & Hansen, J. J. (2008b). Personalizing the Implicit Association Test increases explicit evaluation of target concepts. *European Journal of Psychological Assessment, 24*, 226–236.

Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology, 54*, 14–29.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413–417.

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. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653–667.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598–609.

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge?: The role of structural fit. *Journal of Personality and Social Psychology, 94*, 16–31.

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

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*, 380–391.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 132*, 416–442.

Rogers, R., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*, 207–231.

Rothermund, K., Teige-Mocigemba, S., Gast, A., & Wentura, D. (2009). Minimizing the influence of recoding processes in the Implicit Association Test: The Recoding-Free Implicit Association Test (IAT-RF). *Quarterly Journal of Experimental Psychology, 62*, 84–98.

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 (IAT): Dissociating salience from associations. *Journal of Experimental Psychology: General, 133*, 139–165.

Rothermund, K., Wentura, D., & De Houwer, J. (2005). Validity of the salience asymmetry account of the IAT: Reply to Greenwald, Nosek, Banaji, and Klauer (2005). *Journal of Experimental Psychology: General, 134*, 426–430.

Rudolph, A., Schröder-Abé, M., Schütz, A., Gregg, A. P., & Sedikides, C. (2008). Through a glass, less darkly?: Reassessing convergent and discriminant validity in measures of implicit self-esteem. *European Journal of Psychological Assessment, 24*, 273–281.

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.

Schmukle, S. C., & Egloff, B. (2005). A latent state-trait analysis of implicit and explicit personality measures. *European Journal of Psychological Assessment, 21*, 100–107.

Schnabel, K., Asendorpf, J. B., & Greenwald, A. G. (2008). Using Implicit Association Tests for the assessment of implicit personality self-concept. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The Sage handbook of personality theory and assessment: Vol. 2. Personality measurement and testing* (pp. 508–528). Thousand Oaks, CA: Sage.

Schnabel, K., Banse, R., & Asendorpf, J. B. (2006a). Assessment of implicit personality self-concept using the Implicit Association Test (IAT): Concurrent assessment of anxiousness and angriness. *British Journal of Social Psychology, 45*, 373–396.

Schnabel, K., Banse, R., & Asendorpf, J. B. (2006b). Employing automatic approach and avoidance tendencies for the assessment of implicit personality self-concept: The Implicit Association Procedure (IAP). *Experimental Psychology, 53*, 69–76.

Schröder-Abé, M., Rudolph, A., & Schütz, A. (2007). High implicit self-esteem is not necessarily advantageous: Discrepancies between explicit and implicit self-esteem and their relationship with anger expression and psychological health. *European Journal of Personality, 21*, 319–339.

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition, 25*, 638–656.

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review, 115*, 314–335.

Sherman, S. J., Presson, C. C., Chassin, L., Rose, J. S., & Koch, K. (2003). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology, 22*, 13–39.

Spruyt, A., Hermans, D., De Houwer, J., Vandekerckhove, J., & Eelen, P. (2007). On the predictive validity of indirect attitude measures: Prediction of consumer choice behavior on the basis of affective priming in the picture–picture naming task. *Journal of Experimental Social Psychology, 43*, 599–610.

Sriram, N., & Greenwald, A. G. (2009). The brief Implicit Association Test. *Experimental Psychology, 56*, 283–294.

Stahl, C. (2006). Software for generating psychological experiments. *Experimental Psychology, 53*, 218–232.

Stahl, C., & Klauer, K. C. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods, 39*, 267–273.

Steffens, M. C. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology, 51*, 165–179.

Steffens, M. C., Kirschbaum, M., & Glados, P. (2008). Avoiding stimulus confounds in Implicit Association Tests by using the concepts as stimuli. *British Journal of Social Psychology, 47*, 217–243.

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.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220–247.

Strayer, D. L., & Kramer, A. F. (1994). Strategies and automaticity: I. Basic findings and conceptual framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 318–341.

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. (2007). Evaluating implicit spider fear associations using the Go/No-Go Association Task. *Journal of Behavior Therapy and Experimental Psychiatry, 38*, 156–167.

Teige, S., Schnabel, K., Banse, R., & Asendorpf, J. B. (2004). Assessment of multiple implicit self-concept dimensions using the Extrinsic Affective Simon Task (EAST). *European Journal of Personality, 18*, 495–520.

Teige-Mocigemba, S., Klauer, K. C., & Rothermund, K. (2008). Minimizing method-specific variance in the IAT: A single block IAT. *European Journal of Psychological Assessment, 24*, 237–245.

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavioral Research Methods, 39*, 767–775.

Wentura, D., & Rothermund, K. (2007). Paradigms we live by: A plea for more basic research on the Implicit Association Test. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 195–215). New York: Guilford Press.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101–126.

Wittenbrink, B., & Schwarz, N. (Eds.). (2007). *Implicit measures of attitudes*. New York: Guilford Press.