

Using Implicit Association Tests for the Assessment of
Implicit Personality Self-Concept

Konrad Schnabel

Humboldt-Universität zu Berlin

Jens B. Asendorpf

Humboldt-Universität zu Berlin

Anthony G. Greenwald

University of Washington

Draft of April 26, 2007

Chapter in:

Boyle, G. J., Matthews, G., & Saklofske, H. (Eds.) (2007). *Handbook of Personality Theory and Testing*. London: Sage.

Psychologists have long suspected that people do not have good access to their own thoughts and feelings and that self-exploration is subject to introspective limits. Empirical evidence supports this view (e.g., Nisbett & Wilson, 1977). What results is a compelling claim for measurement procedures suitable for the assessment of cognitive processes that remain obscured if people are simply asked to report them. Recently, Social Cognition research has made progress towards the development of such “implicit” procedures. This chapter deals with a most prominent class of implicit procedures, the Implicit Association Tests (IATs, Greenwald, McGhee, & Schwartz, 1998), and reviews theoretical and empirical contributions that explored IATs for the assessment of the personality self-concept.

Although the term “implicit” was criticized because it left unclear whether it described the measure or the construct, we retain this term that is now firmly entrenched in the literature. Originally, “implicit” was intended to be a label for the measure. In this chapter, we consider both aspects and refer to the implicit personality self-concept at the construct level and to features of the measurement procedure (e.g., unawareness or fakability) at the empirical level. Also, we refer to IAT measures in plural to indicate that they represent different applications of a general procedure rather than one specific test.

The chapter is organized into four main parts. The first section is devoted to the conceptualization of the explicit and the implicit personality self-concept. Following this, we describe essential characteristics of IAT procedures and discuss different theoretical accounts for IAT effects. The third section presents the psychometric properties of IATs including their reliability and their ability to predict criterion variance unpredicted by self-report measures (incremental validity). The section also discusses the extent to which IATs meet the demands of individual diagnosis. In the final section, we describe alternatives to IAT measures along with their advantages and disadvantages over the IATs and offer an outlook to future research.

Implicit and Explicit Personality Self-Concept

Bearing in mind the mind's limited ability to introspect, current Social Cognition research developed *two-systems models* that differentiate between two ways of human information processing. In the following, we employ the Reflective-Impulsive Model (Strack & Deutsch, 2004) to illustrate this approach although it remains debatable whether an implicit-explicit distinction strictly at the empirical level would be sufficient (Greenwald & Nosek, in press). Strack and Deutsch's model proposes that perception, thinking, and behavior are functions of two different systems of information processing, the Reflective and the Impulsive System. In the *Reflective System*, behavior is the result of propositional reasoning. For instance, thinking about one's life may lead to the decision "I am happy". In this reasoning process, information is made available in the form of propositions. Propositions consist of concepts (e.g., 'I' and 'happy') that are linked by a relation (e.g., 'am'). Propositions are generated through introspection and may be considered as either true or false. In the *Impulsive System*, information is processed by the spread of activation between concepts that are associatively linked (e.g., 'I'-'happy'). Associative links are activated spontaneously and are only indirectly accessible by introspection. They may vary in strength but they cannot be considered as true or false.

Dissociations between propositional and associative information processing are also relevant with respect to representations of one's own self. The entirety of one's self-representations is called the self-concept. The part of the self-concept that refers to personality describing attributes is named the personality self-concept (cf. Asendorpf, Banse, & Mücke, 2002). Considering the conceptualization of two-system models, the personality self-concept comprises both propositional and associative representations of one's personal characteristics.

We call propositional representations of the personality self-concept the *explicit personality self-concept* and associative representations the *implicit personality self-concept*. The explicit personality self-concept consists of propositional categorizations that include the concepts of the self and personality describing attributes. The implicit personality self-concept consists of associative clusters between concepts of the self and personality describing attributes.

Although this dual-representation interpretation remains debatable (Greenwald & Nosek, in press) the differentiation between the explicit and the implicit personality self-concept may be useful to conceptualize implications for personality assessment. Given that the explicit personality self-concept contains conscious representations that are generated through a deliberate reasoning process, these representations correspond to answers in a questionnaire that asks respondents to inform about themselves. Consequently, these answers are subject to the limitations of explicit representations (Greenwald & Banaji, 1995). One limitation is that answers in a questionnaire rely on verbal report that is intentionally given to inform about the self. Therefore, questionnaire answers are susceptible to *self-presentational biases*. Second, questionnaire answers only refer to representations of the personality self-concept that are accessible through introspection. Therefore, they are bound to *introspective limits* and may not reflect the entirety of an individual's implicit knowledge about his or her personality. Due to these limitations, there is large interest in the field of personality assessment to have access to procedures that are freed of these limits of explicit questionnaire measures and that are suitable for the assessment of implicit self-representations. Progress has been made with the development of these implicit measures, especially with the development of chronometric procedures like the IAT (Greenwald et al., 1998) that will be described in the next section in more detail.

IAT Procedures

IAT measures are designed to assess automatic associations between a contrasted pair of target (such as ‘me’ versus ‘others’) and attribute (such as ‘anxious’ versus ‘confident’) concepts through a series of discrimination tasks that require fast responding. Faster responses are expected when two highly associated categories (e.g., ‘me’ and ‘confident’) share the same response. IATs start by introducing participants to the target, and, subsequently, to the attribute concept. For instance, an IAT that assesses the personality self-concept of anxiousness¹ (see Table 1) *first* trains participants to press the left response key when a ‘me’ word is presented on the screen and the right response key when an ‘others’ word is presented (side assignments are arbitrary and may be counterbalanced, see below). In the *second* block, participants are trained to press left for ‘anxious’ words and right for ‘confident’ words. The *third* and *fourth* block combine the target and the attribute discrimination, and ask participants to respond left to ‘me’ or ‘anxious’ words, and right to ‘others’ or ‘confident’ words. The combined tasks are subdivided into one block of 20 trials and one block of 40 trials in standard IATs mostly due to historical reasons. Early IAT procedures labeled the first 20 trials as practice blocks and the following 40 trials as test blocks. Currently used IAT procedures do not use practice and test instructions, and scoring algorithms also include data from the first 20 trials because this was shown to increase the IATs’ psychometric properties (Greenwald, Nosek, & Banaji, 2003). Nevertheless, the 20+40 subdivision is often retained and may help to reduce task demands in the otherwise overly lengthy combined blocks. The subsequent *fifth* block reverses the target discrimination, and assigns the left response to ‘others’ words and the right response to ‘me’ words. Finally, the *sixth* and *seventh* block combine again the attribute and the previously reversed target discrimination, and asks participants to respond left to ‘others’ or ‘anxious’ words, and right to ‘me’ names or ‘confident’ words.

- Insert Table 1 here. -

The standard IAT procedure (cf. Nosek, Greenwald, & Banaji, in press) (a) instructs participants to “respond rapidly while occasional errors are acceptable”, (b) displays category labels assigned to the right or left response key in the right or left upper screen corner throughout all tasks, (c) presents labels and stimuli of the target concept in a font (color or type) distinct from the attribute concept when both are represented by printed word stimuli, (d) alternates between target and attribute stimuli in the combined blocks, and (e) employs 60 trials (trial numbers can vary somewhat in typical use) in the combined blocks (each divided into an initial block of 20 trials and a main block of 40 trials), 20 trials in the first two simple discrimination blocks, and 40 trials in the third block (reversed target discrimination). For this block, 40 instead of only 20 trials are recommended because this was shown to reduce the well known order effect for combined blocks (Nosek, Greenwald, & Banaji, 2005). Due to this order effect, mean IAT scores tend to show slightly stronger associations corresponding to the pairings of the combined block that is completed first. To control for this effect, it is often recommended to counterbalance the order of combined blocks between participants (e.g., Nosek et al., in press).

However, the opinions are mixed concerning the experimental variation of procedural variables. When IATs are employed to assess the personality self-concept, and individual differences rather than cross-group differences are the matter of interest, it seems reasonable to keep procedural variables constant. Otherwise, procedural variance is confounded with interindividual variance (cf. Banse, Seise, & Zerbes, 2001). Many procedural variables do not significantly affect IAT scores and their correlations (cf. Nosek et al., in press, for a more detailed discussion), and may be set to current standards (usually five but at least two stimulus items per category, response-stimulus interval of 250 ms, fixed response key assignments).

According to the same logic, the stimuli may be presented in a fixed random order – while alternating target and attribute trials – for all participants in correlational studies (cf. Schnabel, Banse, & Asendorpf, 2006a).

In contrast, there are two procedural variables, namely order of combined tasks and order of implicit-explicit measures, that many favor counterbalancing because they may have effects on IAT results. Experimental variation allows for the statistical compensation of these effects in regression designs and maximizes the generalizability of results. A recent meta-analysis revealed that correlations between IATs and explicit questionnaire measures are higher if the order of compatible and incompatible pairing is counterbalanced across participants (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). There seems to be a better chance to have participants completing the IAT tasks in the order that is optimal for the valid assessment of their characteristic values if one counterbalances the sequence of the tasks. On the contrary, little or no effect on explicit-implicit correlations was found for the order of explicit self-report and IAT measures (Hofmann, Gawronski et al., 2005; Nosek et al., 2005).

At present, order effects as well as their consequences for implicit-explicit correlations are not fully understood. It is quite plausible that answering explicit self-report measures has an effect on IAT responses, especially if the implicit representations are weak or generated on the basis of conscious reasoning (cf. Gregg, Seibt, & Banaji, 2006). On the other side, it was shown that the incompatible block of an flower-insect attitude IAT (i.e., the block that combines ‘flower’ + ‘negative’ and ‘insect’ + ‘positive’) increases response latencies during explicit ratings of attitudes towards flowers and insects while it had no effect on the means of the ratings themselves (Klauer & Mierke, 2005). In conclusion, counterbalancing of order of combined tasks and implicit-explicit measures may be an option in the absence of specific reasons for do otherwise especially in studies where the subsamples of different conditions are

large enough to compare for correlational differences (at least $n = 50$, better $n = 100$, for each subsample).

The different opinions concerning variation of procedural variables reflect different research traditions in experimental and correlational psychology (see already Cronbach, 1957).

Experimental psychology often focuses on the internal validity of experimental procedures and considers individual differences as error variance. In contrast, correlational psychology is interested in the assessment of individual differences and rather considers procedural variance as unwanted error variance. While standards that are established in one tradition are often ignored in the other, both approaches can learn from each other. The most frequently ignored standard from correlational psychology is that correlational designs require adequate sample sizes. It should be noted that the 95% confidence interval for a correlation of .30 ranges from .11 to .47 for $N = 100$, and already from .02 to .53 for $N = 50$. This questions the interpretability of correlations and correlational differences in studies with small samples (N considerably below 50, often even below 30) that are nevertheless frequently published even in first tier journals.

Calculation of IAT scores

For the calculation of IAT scores (IAT effects), only the combined tasks are relevant. Scores are based upon the difference in mean response latencies in the second minus the first combined task (see Table 1). Thus, if participants are quicker in combining ‘me’ + ‘anxious’ and ‘others’ + ‘confident’ relatively to the reverse pairing, they attain low latencies in the first combined task and high latencies in the second combined task. This would result in a positive score in the anxiousness IAT. Consistent with the IAT’s logic (Greenwald et al., 1998),

positive scores in this case reflect *stronger associations* for ‘me’ + ‘anxious’ and ‘others’ + ‘confident’ *relatively* to ‘me’ + ‘confident’ and ‘others’ + ‘anxious’.

- Insert Table 2 here. -

As a refinement to this difference score, Greenwald, Nosek, and Banaji (2003) proposed an improved scoring algorithm that produces an interrelated set of *D measures*. These were shown to increase internal consistencies, correlations with self-report measures, and resistance to the influence of extraneous factors (e.g., general speed of responding). Their major feature is that they are individually calibrated by each respondent’s standard deviation of response latencies. In detail, the algorithm for D measures includes the steps that are described in Table 2. If the IAT procedure does not prompt participants to correct errant responses (i.e., the standard procedure) and rather shows an error message (for, e.g., 800 ms) without the possibility to answer before continuing with the next trial, the D600 or D2SD measures are recommended. These require an extra step after Step 2 and replace latencies of error trials by the block mean for correct trials plus an error penalty of either 600 ms or twice the standard deviation of the same block’s correct responses (cf. Footnote 2, Nosek et al., in press). Prompting participants to correct errant responses saves them to simply wait during presentation of the error message and trains the correct use of response keys. Not requiring error correction allows participants to rush through the test by randomly pressing the right or left key. The latter may be controlled by excluding data of participants with high error rates. Given the lack of empirical evidence comparing these two procedural variations, we recommend to prompt participants to correct errant trials because this is the design that is used most often and that will increase comparability between different studies.

Calculation of internal consistencies

There exist various ways for calculating internal consistencies of IAT D measures. Some compute difference scores for every single trial of the combined blocks and treat them as separate items to calculate Cronbach's internal consistency alpha, some employ difference scores for blocks of 5, 10, 20 or more trials, some calculate split-half reliabilities over blocks with identical number of trials, and some over blocks with different numbers of trials. In order to control for response changes during completion of the IAT, odd-even like estimates may be more useful than block-wise estimates (cf. Schmukle & Egloff, 2006). Procedures that calculate internal consistency estimates over more than two sub-blocks may result in slightly higher scores, although this has not been systematically investigated. The standard procedure calculates split-half reliabilities over the difference scores of Block 6/3 and Block 7/4 (see Table 1).

Accounts for the IAT Effect

IATs operate on the basic premise that responses are easier (i.e., quicker) if the concepts assigned to the same response key are strongly associated (Greenwald et al., 1998). Although there does not yet exist a comprehensive theoretical account, that explains underlying psychological mechanisms of this "IAT effect" and elucidates method-specific influences, the process of developing such an account has been begun in several publications.

De Houwer's (2003a) approach referred to a *stimulus-response compatibility* mechanism and claimed that in the two combined blocks stimuli elicit either compatible or incompatible response tendencies. Whereas responses are unambiguously associated with a certain meaning in the block that is completed faster, response representations are more ambiguous in the

block that is completed more slowly. The combined block that is completed faster is often referred to as the “compatible” IAT block. Of course, compatibility is a function of the participant, and blocks should rather be labeled as the, e.g., “me-anxious” or “me-confident” pairing. For didactical purposes, De Houwer’s stimulus-response compatibility account is explained using the flower-insect IAT because this IAT has a clear compatible (‘flower’ + ‘positive’ and ‘insect’ + ‘negative’) and incompatible (‘flower’ + ‘negative’ and ‘insect’ + ‘positive’) pairing. De Houwer proposes that responses to items of the target concept may be based on either their category membership (e.g., ‘flower’ versus ‘insect’) or their evaluative meaning (e.g., ‘positive’ versus ‘negative’). Both the category-based and the valence-based response tendencies lead to correct responses in the compatible pairing (synergistic effect). In the incompatible pairing, these response tendencies interfere with each other and only the category-based response tendencies lead to correct responses in terms of key assignments (antagonistic effect).

Steffens and Plewe (2001) claimed that both a concept-based and a stimulus-based factor have an impact on the IAT effect. The concept-based factor refers to target-attribute associations at the concept level. It accounts for a simplified task representation during the compatible IAT task due to a dimensional overlap (e.g. “good-bad”) between the target (e.g. “flower-insect”) and the attribute (e.g., “positive-negative”) concept. The stimulus-based factor refers to individual features of target and attribute stimuli. It accounts for a modification of the concept effect depending on whether cross-category associations at the stimulus level are consistent (e.g., insect = cockroach, wasp) or inconsistent (e.g., insect = ladybird, firefly) with associations at the concept level. Steffens and Plewe (2001) manipulated cross-category associations in a gender IAT and showed that the IAT effect was larger for item sets with consistent rather than inconsistent cross-category associations. However, inconsistent cross-category associations did not reverse the IAT effect. Steffens and Plewe (2001) concluded

that the concept factor, plays a major role in the IAT effect but may be substantially altered by stimulus features (for similar results, cf. Mitchell, Nosek, & Banaji, 2003; Nosek, Greenwald, & Banaji, 2005; Rudman, Greenwald, & McGhee, 2001; for cross-category associations reversing the IAT effect, cf. Blümke & Friese, 2003; Govan & Williams, 2004). Concerning practical implications, these findings suggest to thoroughly select IAT stimuli in a way that they (a) well represent all relevant aspects of the concept category and (b) may not be categorized according to features different from the category frame (Nosek et al., 2005).

Additional evidence for the key influence of the task factor on the IAT effect was presented by Olson and Fazio (2005). These authors developed personalized IAT variants in order to assess attitudes towards various targets (e.g., apple versus candy bars, Bush versus Gore). Simply by changing the category labels of the attribute concept from ‘positive’ and ‘negative’ to ‘I like’ and ‘I dislike’, they found higher implicit-explicit correlations for these personalized IATs than for standard IATs. The authors concluded that the personalized IATs reduced effects of extrapersonal associations and focused the IATs on more personal associations.

Extrapersonal or environmental associations (cf. Karpinski & Hilton, 2001) were assumed to reflect external estimations that one has encountered in society (e.g., “Apples are healthier and should be judged more positively than candy bars.”) and that do not necessarily correspond with one’s personal attitudes (e.g., “Sometimes, I really like to have a candy bar.”). It may be worthwhile to adapt self-concept IATs to this personalized form in order to reduce unwanted environmental effects. This adaptation may not be easy because many self-concept IATs do not employ an evaluative attribute concept and may not be adapted using Olson and Fazio’s ‘I like’ and ‘I dislike’ category labels. Additionally, as a first step, empirical evidence should be collected that shows that personal IATs are superior to standard

IATs in the prediction of social behavior that is in turn a function of environmental and contextual influences. Recent studies suggest that the personalized IAT does not remove a confound of extrapersonal variation. Instead, the ‘I like’ and ‘I dislike’ labels for the target categories rather introduce a confound of task-recoding in which the target stimuli are more likely to be explicitly evaluated (Nosek & Hansen, 2007).

Mierke and Klauer (2001; 2003) reported a *task-switching* account of the IAT effect. Similarly to De Houwer (2003a), they stated that merely considering attribute-related information is sufficient for fast and accurate responding within the compatible IAT task. Consequently, participants may neglect to switch between target-based and attribute-based decisions in the compatible pairing. As participants neglect to switch they avoid task-switching costs that are inescapable in the incompatible task. Mierke and Klauer (2001) compared response latencies between trials that switched between target to attribute discrimination and trials that required a discrimination according to the concept of the preceding trial. Results showed that switching between target and attribute discrimination produced significantly more costs (i.e., longer response latencies) in the incompatible than in the compatible IAT task.

Another set of experiments showed that interindividual differences in task-switching performance produce reliable method-specific variance in IAT scores, although the improved scoring procedure (see above) seems to control for this effect. Mierke and Klauer (2003) demonstrated that IAT effects could be obtained with an IAT that was not based on pre-existing associations between target and attribute concept. For this purpose, they developed an IAT that experimentally imposed a contingency between the target features (‘blue’ versus ‘red’) and the attribute features (‘big’ versus ‘small’) of geometrical objects, in such a way that all blue objects were big and all red objects were small. This geometrical objects IAT

produced IAT effects that were internally consistent (Cronbach's $\alpha = .93$) and showed a significant correlation with the absolute scores of a self-concept IAT that measured extraversion ($r = .29$). The authors used absolute scores for the extraversion IAT because they expected interindividual differences in task-switching performance to primarily affect the incompatible IAT pairing. Whether a pairing is compatible or incompatible, however, is a function of the participant rather than a function of the IAT. Participants with poor task-switching performance decrease response latencies in the pairing that is their particular incompatible pairing (i.e., either the 'me-extraverted' or the 'me-introverted' pairing). Consequently, these participants add an extremity bias to their IAT results rendering them either more extraverted or more introverted. The extremity bias is better captured by absolute scores rather than by IAT raw scores. In contrast, absolute scores and raw scores were identical for the geometrical objects IAT because no participant showed negative scores in this IAT. Overall, there was no conceptual similarity concerning the content of the extraversion IAT and the geometrical objects IAT, and the significant correlation between both IATs could not be interpreted in terms of convergent validity. Thus, Mierke and Klauer's (2003) results indicated a reliable contamination of both IATs with method-specific variance. Interestingly, when Mierke and Klauer (2003) computed the IAT scores as D measures using the improved scoring algorithm presented by Greenwald et al. (2003), the correlation between the geometrical objects IAT and the extraversion IAT was substantially reduced and no longer significant ($r = .12$).

In a similar vein, Back, Schmukle, and Egloff (2004) explored correlations between another IAT measuring task-switching abilities and a content-specific self-concept IAT measuring anxiousness. Their task-switching IAT differed from Mierke and Klauer's (2003) IAT in two aspects that aimed to improve structural closeness to features of content-specific IATs. First, Back et al.'s IAT employed verbal material instead of geometrical objects, and the concept

categories were ‘letter’ (e.g., M, B) versus ‘number’ (e.g., 4, 7) and ‘word’ (e.g., shirt, pen) versus ‘calculation’ (e.g., $7 - 4 = 3$, $4 + 5 = 9$). Second, it used the preexisting associations between these concept categories (i.e., between ‘letter’ and ‘word’, and between ‘number’ and ‘calculation’) instead of experimentally imposed contingencies between geometrical objects in order to avoid individual differences caused by learning the previously unknown associations.

Similar to the results by Mierke and Klauer, Back et al.’s results showed that correlations between the task-switching and the anxiousness IAT can be significantly reduced if the improved D measures were used for the IATs (in a combined analysis of three studies, from $r = .29$ to $r = .17$). Differently from Mierke and Klauer, the remaining small correlation between the two conceptually unrelated IATs was still significant in Back et al.’s study. This may be attributed to the larger sample size in Back et al.’s study and the effect that their task-switching IAT may be more appropriate to capture task-switching costs in content-specific IATs. Together, these findings suggest that task-switching costs (greater costs when categories sharing a response key are not well associated) represent an important component of the IAT effect. Additionally, a residual small portion of the IAT measure seems to contain a component of task-switching ability that is independent of the association strengths being measured.

Rothermund and Wentura (2004) suggested a *figure-ground model* that posits that target and attribute categories are asymmetrical with respect to their salience. According to this model, the salient category of the target (e.g., ‘Blacks’) and the attribute (e.g., ‘negative’) concept serves as “figure” on the “ground” of the opposing nonsalient categories (e.g., ‘Whites’ and ‘positive’). When the salient categories are mapped to one response key and the nonsalient categories are mapped to the other response key during the compatible pairing, participants

can base their discrimination on the figure-ground asymmetries alone. In contrast, there is a mismatch of “figure” and “ground” categories in the incompatible pairing.

In a series of different experiments, Rothermund and Wentura (2004) dissociated effects of salience asymmetries from effects of associations between target and attribute concepts, and showed that salience asymmetries alone may produce IAT effects. However, this does not rule out that differences in association strength may also be a source of IAT effects, and that salience asymmetries themselves may be the result of such differences. For example, insects may be the salient category in the flower-insect IAT because insects are associated with negative valence. As a consequence, salience asymmetries may only distort IAT effects in artificial IATs that were designed to maximize salience asymmetries and to minimize associations between concept categories (Greenwald, Nosek, Banaji, & Klauer, 2005). Furthermore, results for the content-unspecific task-switching IATs (Back et al., 2004; Mierke & Klauer, 2003) imply that salience asymmetries are, like associations, not a necessary precondition for IAT effects. Thus, the salience asymmetry account has not yet established itself as an aid to understanding the mechanism of IAT effects.

A key role in many of the accounts listed before seems to be whether target and attribute categories share similar features (cf. De Houwer, Geldof, & De Bruycker, 2005). The more similarity exists between target and attribute concepts, the more similar are the activation patterns that they produce. When similar activation patterns are matched to identical responses in the compatible IAT pairing, responses are facilitated. The origins of similarity may be manifold, and shared salience asymmetries may be just one source of similarity. In turn, features that are shared between different concepts may provide a basis for associations between concepts. If one is interested in the assessment of associations one has to consider whether the similarity between the concepts refers to the associations of interest rather than to

alternate features of shared similarity (e.g., word length, stimulus familiarity, salience asymmetries; cf. Dasgupta, McGhee, Greenwald, & Banaji, 2000; Greenwald et al., 2005).

It is important to note that most of the accounts that aim to explain IAT effects refer to mechanisms that affect mean IAT effects. Only a few studies refer to effects on correlations between different IATs or between IATs and explicit measures. Notably, correlations between IATs and explicit measures may be unaffected even if IAT effects are reduced by procedural variations or stimulus features (e.g., Nosek et al., 2005; Steffens & Plewe, 2001). Only one account explicitly refers to an individual differences perspective and suggests that task-switching abilities constitute a (small) contaminant of IAT measures (Back et al., 2004; Mierke & Klauer, 2003). Generally, care should be taken if models concerning mean IAT effects are employed to draw conclusions about the correlations of IAT scores.

Psychometric Properties of IATs

Evaluating whether IATs meet relevant psychometric criteria is especially important when IAT measures are employed for the assessment of individual differences. A proper assessment requires that IATs refer to relatively stable individual differences in implicit self-representations, and that these differences contribute significantly to the prediction of behavior. In this section, we review psychometric properties of IAT measures and refer particularly to IAT adaptations that deal with the assessment of the implicit personality self-concept.

Reliability

Various adaptations of IAT measures usually reach *internal consistency* estimates (split-half correlations or Cronbach's alphas) between .70 and .90 (e.g., Banse et al., 2001; Schmukle & Egloff, 2004). Such reliabilities are psychometrically satisfactory, and they are much higher than those found for other latency-based measures such as priming procedures (e.g., Bosson, Swann, & Pennebaker, 2000; Kawakami & Dovidio, 2001) the Go/No-Go Association Task (GNAT, Nosek & Banaji, 2001) or the dot probe task (Schmukle, 2005). Although internal consistencies for IATs may depend on the method of calculation (see above), these influences have not yet been studied systematically and may be relatively minor.

Another measure of reliability, test-retest reliability, has been observed to show a median of .56 across different studies (Nosek et al., in press) which is about .15 to .20 below the internal consistencies that are typically obtained for IAT measures (Greenwald & Farnham, 2000; Schmukle & Egloff, 2004; Schnabel, Banse, & Asendorpf, 2006b). The same is true if the retest is completed after a time span of up to one year (Egloff, Schwerdtfeger, & Schmukle, 2005), or if the second test is a parallel test that employs parallel attribute stimuli (Asendorpf et al., 2002). Within the period of one year it seems rather irrelevant whether the retest is completed immediately or with relatively more time after the first test (cf. Nosek et al., in press). The reasons for the discrepancy between a satisfactory internal consistency and a retest reliability that is somewhat too low for assessments of stable constructs are presently unclear. The lower test-retest reliability implies that there are systematic occasion-to-occasion variations in IAT scores that stem from changes in either (a) the association strengths being measured or (b) unidentified additional sources of variance in the measurement procedure. The first is plausible because of the numerous context effects that have been demonstrated for IAT measures (Blair, 2002).

Approaches that separated trait and state influences (Schmukle & Egloff, 2004, 2005) showed that IATs capture both reliable trait-specific and state-specific variation. However, state-specific variation in an anxiousness IAT could, in contrast to an increase in self-reported state-anxiety, not be explained by an anxiety induction (Schmukle & Egloff, 2004).

Importantly, Schmukle and Egloff (2004) found non-significant induction effects on the anxiousness IAT both when using a between-subjects design (IAT was presented after the anxiety induction and compared to a control group without anxiety induction) and a within subjects-design (IAT was presented before and after the anxiety induction). The usage of between-subjects designs for the exploration of state effects on IAT measures is indispensable due to the small but systematic attenuation of IAT effects from first to subsequent administrations (Greenwald et al., 2003). One option for repeated measures designs may be to include a nonrelevant control IAT that is expected to be unaltered by the manipulation and to compare effects on the relevant IAT with effects on the control IAT (cf. Teachman & Woody, 2003). But even then, there remains some uncertainty about the comparability of repeated administration effects on the relevant and the control IAT.

Fakability

Investigations of the fakability of IATs revealed that they are, though slightly fakeable, much less fakeable than explicit self-reports (Asendorpf et al., 2002; Banse et al., 2001; Boysen, Vogel, & Madon, 2006; Egloff & Schmukle, 2002; Steffens, 2004). Fakability increased if participants were informed beforehand about how to fake (Kim, 2003). Importantly, faking effects on mean IAT scores are a threat to the validity of individual differences measured by IATs only if differential faking (different individuals fake to a different extent) occurs.

Differential faking effects should change the rank order of participants' IAT scores and alter their correlations with external validation criteria that are expected to be immune to faking.

Schnabel and colleagues (2006b) explored differential faking effects on a shyness IAT and a parallel chronometric procedure, the shyness IAP, and revealed strong effects of differential faking for explicit self-report but not for implicit measures. The authors compared correlations between an experimental group that was instructed to appear non-shy and a control group without faking instructions. Results showed that (a) implicit-explicit correlations were moderate in the control group whereas they were significantly reduced in the experimental group, (b) faking instructions increased the correlation between explicit shyness and a social desirability scale whereas the correlation between implicit shyness and social desirability was low in both groups, and (c) faking instructions decreased the correlations with observer judgments of shyness for explicit but not for implicit measures of shyness. Thus, IATs seem to be more robust against faking attempts than explicit self-report measures with regard to both mean and differential faking effects.

The fact that IATs are slightly fakeable suggests that they are not process pure in the sense of measuring solely the strength of automatic associations. A recent multinomial modeling approach differentiates between several automatic and controlled cognitive processes in implicit social cognition and provides evidence that IATs may also reflect controlled efforts to reduce automatic biases (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005). Future research should deal with the question of how making IAT measures less fakeable. Another approach may be to develop algorithms that allow to empirically distinguish honest from faked IAT performance (Cvencek & Greenwald, 2006).

Validity

Convergent and discriminant validity with implicit measures. Correlations of IATs with other implicit measures are typically weak. Bosson, Swann, and Pennebaker (2000) observed nonsignificant correlations between a self-esteem IAT and six other implicit self-esteem measures. Correlations between IATs and priming procedures tend to be small or nonsignificant (Olson & Fazio, 2003). For the assessment of individual differences in implicit self-representations, priming procedures are scarcely used mainly due to their unsatisfactory reliability (cf. Banse, 1999). If one accounts for the lack of reliability, the observed disattenuated correlations between priming procedures and IATs become more substantial (Cunningham, Preacher, & Banaji, 2001). Insufficient reliability of other implicit measures may principally explain why their correlations with IAT measures are so small. This is true also for recently developed implicit procedures (for a description of these measures, see next section) like the GNAT (Nosek & Banaji, 2001) and the EAST (De Houwer & De Bruycker, in press; Teige, Schnabel, Banse, & Asendorpf, 2004).

For a different new implicit procedure, the Implicit Association Procedure (IAP), correlations with an IAT were much higher (up to .50) and not much lower than the IAT's and the IAP's retest reliability (Schnabel et al., 2006b). The IAP procedure is similar to the IAT in that it measures relative association strengths by comparing response latencies of two combined discrimination tasks. Differently from the IAT, the IAP uses joystick movements towards or away from the participant instead of a right and left response key. Because the IAP is similar to the IAT in other methodological details, the correlation of approximately .50 between IAT and IAP found by Schnabel et al. (2006b) may indicate an upper bound to the construct validity, suggesting a substantial method variance in the IAT and similar procedures.

Convergent and discriminant validity with explicit self-report measures. In several fairly large samples ($N > 98$) Asendorpf and colleagues found small to moderate correlations between self-concept IATs and explicit self-ratings on IATs' attribute stimuli. For a shyness IAT, correlations ranged between .30 and .44 (Asendorpf et al., 2002; Schnabel et al., 2006b; Teige et al., 2004), for an angriness IAT between .11 and .39 (Schnabel, Banse, & Asendorpf, 2006a; Teige et al., 2004), and for an anxiousness IAT the implicit-explicit correlation was .25 (Schnabel et al., 2006a). The small correlation for the anxiousness IAT replicated results from studies by Egloff and colleagues who used 'anxious' versus 'calm' instead of 'anxious' versus 'confident' as attribute categories (Egloff & Schmukle, 2002, 2003). Steffens and Schulze-König (2006) explored implicit-explicit correlations for the Big Five in a total sample of $N = 89$ psychology students and found significant but small correlations ($< .29$) for neuroticism and conscientiousness, but nonsignificant correlations for the other Big Five traits (i.e., extraversion, openness, agreeableness). Together, these findings are in line with results from other areas of IAT research (e.g., attitudes, stereotypes) revealing that IATs and their corresponding explicit self-report measures show evidence for both convergent and discriminant validity and refer to related but distinct constructs (Nosek & Smyth, in press).

Recent meta-analyses on correlations between IATs and explicit self-reports over numerous content domains (including attitude, self-concept, and stereotype IATs) revealed average implicit-explicit correlations of .24 (Hofmann, Gawronski et al., 2005) and .37 (Nosek, 2005). The difference between these findings may be explained by the facts that (a) the studies put a focus on different domains and (b) Nosek (2005) used relative feeling thermometers as explicit measures exclusively that may correspond more closely with the IAT in the sense that they more readily tap into an affective component. Hofmann and colleagues (2005) included studies using various explicit measures and content domains that are likely to elicit only weak

implicit-explicit consistency (e.g. racial attitudes) whereas most of Nosek's (2005) studies referred to attitude domains showing moderate to substantial implicit-explicit correlations.

Hofmann and colleagues (2005) organized variables that may moderate implicit-explicit consistency in a process model containing five primary factors. The *translation* factor refers to the interrelation between implicit and explicit representations. It includes aspects like representational strength (subjectively important or frequently processed representations), dimensionality (representations that refer to either end of a bipolar continuum), social distinctiveness (representations thought to be distinct from other individuals), and awareness (introspectively accessible representations) that are all associated with greater implicit-explicit consistency (cf. Nosek, 2005). The factor *additional information integration* describes whether explicit representations assessed by verbal self-report are generated spontaneously or deliberately. Explicit representations that are generated spontaneously or with minimal use of cognitive resources show greater implicit-explicit consistency.

The factors *explicit assessment* and *implicit assessment* refer to reliability and method-specific variance (e.g., fakability) of explicit and implicit measures. With respect to the explicit assessment factor, differences in social desirability or self-presentational concerns may also moderate implicit-explicit consistency (Nosek, 2005), although this influence seems to be somewhat over-estimated (Hofmann, Gschwendner et al., 2005) and has not yet been found for correlations with self-concept IATs (Egloff & Schmukle, 2003). With respect to the implicit assessment factor, the situational malleability of implicit representations (e.g., Blair, 2002) may additionally play a role in decreasing implicit-explicit consistency. Finally, *design factors* such as variance restriction due to sampling biases and lack of conceptual correspondence between implicit and explicit measures were found to reduce implicit-explicit consistency.

Recently, Nosek (in press) provided a nice illustration that the correlational or convergent validity indicator of implicit-explicit consistency can vary independently from particular mean differences between implicit and explicit measures. The latter refer to whether implicit and explicit measures show different deviations of the sample means from theoretically expected zero points on the scales. Using attitude IATs as examples, Nosek showed that even if implicit and explicit preferences were at odds concerning the sample means (e.g., indicating explicit preferences for ‘evolution’ and implicit preferences for ‘creationism’) interindividual correlations of these preferences can be very high (in this case .60). Moreover, there was no systematic relationship across 58 different attitude objects between the consistency of implicit and explicit preferences concerning sample means and the consistency concerning the implicit-explicit correlations.

Predictive validity for behavioral measures. A recent meta-analysis (Poehlman, Uhlmann, Greenwald, & Banaji, 2007) found compelling evidence for the predictive validity of IATs (but also explicit measures) across various behavioral domains. The predictive validity of explicit measures, differently from IAT measures, was reduced in domains that are hardly guided by conscious control or that are susceptible to social desirability concerns. IATs outperformed explicit measures in the domain of stereotyping and prejudice, whereas explicit measures outperformed IATs in studies that explored brand preferences or political candidate preferences. Perugini (2005) classified predictive models of implicit and explicit measures to into three different types: the additive, the multiplicative, and the double dissociation model. All three models postulate that implicit measures show incremental validity and increase the prediction of behavior. This aspect is crucial for judging the value of implicit procedures for the assessment of personality constructs. The *additive model* describes implicit and explicit measures explaining different portions of variance of a relevant criterion. Concerning self-

concept IATs, additive validity of a shyness, an anxiousness, and an angriness IAT was found for the prediction of observer judgments of shy behavior (Schnabel et al., 2006b), anxious behavior (Schnabel et al., 2006a), but not angry behavior (Schnabel et al., 2006a), respectively.

In the *multiplicative model*, implicit and explicit measures interact in predicting relevant behavioral criteria. Interactive validity effects were found for self-esteem IATs indicating that persons with discrepant self-esteem (persons showing discrepancies between implicit and explicit self-esteem) exhibit more defensive behavior (rejection of negative feedback, exaggerated social consensus estimates) than individuals with congruent self-esteem (Jordan, Spencer, Zanna, Hoshino-Browne, Correll, 2003; McGregor, Nail, Marigold, & Kang, 2005; Schröder-Abé, Rudolph, Wiesner, & Schütz, in press). Defensive behavior was explained by a lack of integration in self-representation for individuals with discrepant self-esteem. Interactive predictive validity for implicit and explicit self-esteem may explain why Poehlman and colleagues (2007) found no overall main effects for implicit self-esteem predicting relevant behavioral criteria.

The *double dissociation model* claims that implicit measures predict spontaneous behavior whereas explicit measures predict controlled behavior. Double dissociation models for IATs have rarely been realized probably due to the complexity of identifying relevant indicators of spontaneous and controlled behavior. McConnell and Liebold (2001) presented evidence that a race IAT measuring attitudes toward Blacks versus Whites significantly correlated with indicators of spontaneous behavior indicating negative Black prejudices (e.g., less smiling towards a Black than a White experimenter). In contrast, the explicit measure of prejudice did not correlate with any of these behavioral indicators. Concerning self-concept IATs, Egloff and Schmukle (2002) found that an anxiousness IAT predicted several behavioral indicators

of anxiety during a stressful speech whereas the explicit anxiousness measure only accounted for self-reported state anxiety during the speech. In a similar vein, Steffens and Schulze-König (2006) showed that four of five self-concept IATs measuring the Big Five correlated significantly with relevant indicators of spontaneous behavior, but explicit self-reports of the Big Five did not.

In what seems to be the strongest finding of a double dissociation, Asendorpf and colleagues (2002) confirmed for shyness a full and strong double dissociation pattern. Their studies included valid indicators of both spontaneous and controlled shy behavior, and the IAT uniquely predicted spontaneous but not controlled behavior whereas self-reports uniquely predicted controlled but not spontaneous behavior. Double dissociation designs are highly recommendable to elucidate the construct validity of implicit measures and to demonstrate their specific contribution in the sense of incremental validity to explicit measures. They are the only way to show unique validity of implicit measures for the prediction of spontaneous behavior while simultaneously ensuring that this cannot be attributed to a general lack of validity of the corresponding explicit measure.

It should be noted that all self-concept IATs that were valid for the prediction of behavior employed attribute concepts that are confounded with positive and negative valence (e.g., anxious versus confident, shy versus nonshy, agreeable versus disagreeable). To date, it is unclear to what extent IAT responses are based on the specific semantic meaning of the attribute categories or simply on their positive and negative valence (cf. Schnabel et al., 2006a). For future self-concept IATs, it will be useful to show construct validity for their specific semantic content as distinct from a general positive or negative evaluation. Rudman and colleagues (2001) employed gender IATs to show independent effects of stereotyping and evaluation in implicit gender associations. In a similar vein, Amodio and Devine (2006)

separated stereotyping and evaluation effects in implicit race biases using evaluative (pleasant versus unpleasant) and stereotyping (mental versus physical) race IATs (Amodio & Devine, 2006).

Alternative Methods to IAT Procedures

Alternative methods intended to overcome the limitation that IATs are restricted to the assessment of relative association strengths and that the concepts need to have two categories. In fact, IAT scores reflect the association strength between one pairing of target and attribute categories relatively to the reverse pairing. For instance, a positive score in the standard flower-insect IAT does not indicate that one evaluates flowers positively and insects negatively. The positive score rather reflects that one evaluates flowers more positively than insects. As the most radical alternative to dual-category concepts, Blanton, Jaccard, Gonzales, and Christie (2006) presented the *Single Association Test (SAT)*.

The SAT assesses the association strengths between the concept categories separately in different tasks. In these tasks, one target and one attribute category is used to create one joint category (e.g., “flower-positive”, “flower-negative”, “insect-positive”, or “insect-negative”) that is contrasted to two control categories. Blanton and colleagues (2006) used ‘furniture’ (e.g., table, desk) and ‘middle’ (e.g., midpoint, halfway) as control categories (cf. Pinter & Greenwald, 2005). They expected the associations between these control categories to be constant for all participants. Consequently, individual differences in response latencies should only be influenced by the associations between the relevant concepts and not by associations between the control concepts. Blanton and colleagues explored the SAT using the categories ‘Black’ (African American names), ‘White’ (European American names), ‘positive’ (positively valenced words), and ‘negative’ (negatively valenced words) of the race IAT. Differently from current standards, they did not employ difference scores between the

combined tasks nor scoring algorithms according to Greenwald et al.'s (2003) D measure in order control for interindividual differences in response latencies. Instead, they employed response latencies of a flower-insect IAT as a general response speed measure.

Using structural equation modeling and partialling general processing speed, Blanton and colleagues showed that the SAT data did not fit structural equation models that assumed that positive and negative evaluations of the racial groups were opposed to each other. Thus, response latencies in the 'Blacks' + 'positive' task did not correlate negatively with response latencies in the 'Blacks' + 'negative' task. The same was true for the 'Whites' + 'positive' and 'Whites' and 'negative' tasks. Both would be required by standard IAT procedures that do not allow to assess these associations separately. In contrast, model fit of the SAT data was good, when response latencies in the four different SAT tasks were treated as separate attitude indicators. Additionally, only response latencies in the 'Blacks' + 'negative' task showed a significant relationship with the scores on a racism questionnaire, suggesting that the tendency to automatically associate Blacks with negative attributes is a key predictor of racism.

Unfortunately, Blanton and colleagues did not directly compare their method to control for general speed differences with algorithms that are usually employed for IAT measures (i.e., difference scores and D measures). Also, the study did not directly compare SAT results with results yielded from an IAT. Thus, it is somewhat difficult to estimate to what extent the SAT approach is superior to standard IAT methods. Most importantly, before the SAT will be broadly employed for the assessment of interindividual differences it needs to show satisfactory internal consistencies. Blanton and colleagues do not report any reliability indices. Additionally, Nosek and Sriram (in press) employed structural equation modeling to show that IAT scores represent *relative measures* that contrast performance between two interdependent conditions and that cannot be decomposed into additive combinations of two

distinct indicators of the same construct. Structural equation models that respected the interdependence of the two IAT tasks by using difference scores resulted in very good model fits whereas models that considered the two IAT tasks as independent additive indicators fit poorly to the data.

A less radical approach to tackle the problem that four categories are confounded in standard IATs is presented by *Single Category IATs* (Karpinski & Steinman, 2006; Nosek & Banaji, 2001; Penke, Eichstaedt, & Asendorpf, 2006; Wigboldus, Holland, & van Knippenberg, 2005). Single Category IATs employ one unipolar concept (e.g., 'Blacks') and one concept that includes two categories (e.g., 'positive' versus 'negative'). There is a broad range of concepts that do not have a natural complement. As a consequence, it makes little sense to consider associations with these non-bipolar concepts in relation to another category (e.g., attitudes towards certain brands). Different variants of single category IATs were shown to show satisfactory internal consistencies and higher implicit-explicit correlations than corresponding standard IAT procedures (Karpinski & Steinman, 2006; Penke et al., 2006). For instance, Karpinski and Steinman showed that a self-esteem Single Category IAT (using 'positive' versus 'negative' as attribute categories and 'me' as target category) was significantly correlated with an explicit self-esteem measures whereas a self-esteem IAT (using identical attribute categories and 'me' versus 'others' as target categories) showed non-significant implicit-explicit correlations. Considering self-concept IATs, researchers may also wish to assess associations between the self and personality traits using Single Category IATs because these associations do not need to be inversely related to associations towards other persons.

However, IAT variants using single categories did not show satisfactory reliability indices in all instances (Nosek & Banaji, 2001; Wigboldus et al., 2005). This may be attributed to the

fact, that categorizing stimuli in Single Category IATs may be facilitated by concentrating on the single category. Consider that in these tasks only one response key is assigned to a joint category (e.g., ‘me’ + ‘positive’) whereas the other response key is assigned to a single category (e.g., ‘negative’). If participants base their responses on the single category and ignore the joint category, the task may no longer assess what one aims to measure, namely, associations within the joint categories. A similar logic may apply to Blanton et al.’s (2006) Single Association Test. Remember that in this task the control category, though it is a joint category (i.e., ‘furniture’ + ‘middle’), remains constant throughout the whole task, while the joint categories containing the relevant concepts are changing. Similarly to Single Category IATs, participants may concentrate on the non-relevant control category in order to facilitate their responding. Also, it is not clear whether response windows as suggested by Karpinski and Steinman (2006) really decrease the likelihood that participants engage in such unwanted processing of the stimuli. Obviously, a lot of work needs to be done in order to know the conditions that make these IAT variants superior to the standard IAT procedure.

De Houwer (2003b) proposed the *Extrinsic Affective Simon Task* (EAST). Similarly to IATs, the EAST requires a double discrimination task including two concepts. Differently from IATs, there is only one double discrimination task in the EAST, and only one concept is categorized according to its relevant feature (e.g., ‘positive’ versus ‘negative’) whereas the second concept (e.g. ‘insect’ versus ‘flower’) is categorized according to a non-relevant feature (e.g., ‘green’ versus ‘red’). The task involves that half of the stimuli for both categories of the second concept have to be responded to with the same key as the, e.g., ‘positive’ category, and the other half has to be responded to with the same key as the, e.g., ‘negative’ category. For instance, if the stimuli of the second concept have to be categorized according to their green or red color, half of the words representing the ‘insect’ category are displayed in red and the other half is displayed in green. The same is true for the ‘flower’

category. In contrast, the stimuli of the first category are displayed in black and have to be categorized according to their meaning (e.g., ‘positive’ versus ‘negative’). Differently from the IAT, response key assignments are fixed for the whole task (e.g., ‘positive’ and ‘green’ are assigned to the left and ‘negative’ and ‘red’ are assigned to the right response key) but the stimuli for the categories of the second concept change their non-relevant feature (e.g., green or red color). Separate EAST scores can be calculated by comparing response latencies and error rates for ‘flower’ and ‘insect’ stimuli assigned to the ‘positive’ versus the ‘negative’ key.

In theory, the EAST allows for the assessment of multiple concepts that are each mutually assigned to the two categories of the non-relevant feature. Unfortunately, the internal consistencies of the EAST proved to be relatively weak (De Houwer, 2003b). Although some adaptations of the EAST reliably and validly assessed individual differences in fear of spiders (Huijding & de Jong, 2005; Ellwart, Becker, & Rinck, 2005), EAST adaptations for the assessment of associations between the self-concept and personality attributes showed insufficient reliability and validity and were clearly outperformed by the psychometric properties of corresponding IAT measures (Teige et al., 2004; De Houwer & De Bruycker, in press). Consequently, the EAST does not seem to be promising for the assessment of the implicit personality self-concept.

Payne, Cheng, Govorun, and Steward (2005) developed a variant of affective priming, the *Affect Misattribution Procedure* (AMP). In this task, participants have to categorize neutral Chinese characters as pleasant or unpleasant. The Chinese characters are preceded by positive, negative, or neutral stimuli (i.e., the primes). Results showed that the judgment of the Chinese characters was influenced by the valence of the primes (misattribution effect). Differently from other priming procedures, preference scores (calculated by comparing character judgments following positive versus negative primes) were internally consistent and

substantially correlated with self-reported preference measures. Theoretically, the AMP allows for the assessment of attitudes towards multiple unipolar targets and even for the assessment of single associations in the sense of the SAT (Blanton et al., 2006) if one uses neutral primes as a reference concept. However, none of these variants has been empirically tested so far. Also, it is not clear whether the AMP can be used for the assessment of the personality self-concept by using, e.g., ‘me’ and ‘others’ as primes and asking to guess the semantic meaning of unknown characters as, e.g., ‘anxious’ or ‘confident’.

Recently, Greenwald (2005) proposed an IAT variant, the *Multifactor Trait IATs* (MFT-IATs), to allow for the assessment of multiple constructs and to deal with the problem that evaluative valence and specific semantic meaning are confounded in many self-concept IATs (cf. Schnabel et al., 2006a). The procedure employs attribute categories that are valence-matched (i.e., all attributes are either positive, negative, or neutral) to avoid category discriminations simply on the basis of their positive and negative valence. MFT-IATs require, similarly to standard IATs, one IAT per attribute concept. Different from standard IATs, attribute concepts in MFT-IATs have only one relevant category referring to the concept and one irrelevant category referring to the other attributes of the MFT-IAT.

For instance, a Big Five MFT-IAT assesses automatic associations between the concept of self and extraversion by using ‘me’ versus ‘others’ as target concept and ‘extraversion’ (e.g., energetic, bold, active, gregarious) versus ‘other trait’ (e.g., agreeableness, conscientiousness, emotional stability, openness) as attribute concept. The same logic applies for the remaining four Big Five traits. It is not yet clear, however, how many trials in the combined blocks are necessary for the reliable and valid assessment of traits using an MFT-IAT. Also, the procedure is relatively demanding because it requires one IAT for each trait of interest.

Positive aspects are that MFT-IATs allow for (a) using single category concepts and (b) balancing the valence of these concepts. Valence-matched concept categories are otherwise difficult to achieve because many dual-category concepts (e.g., anxious versus calm, shy versus nonshy, agreeable versus disagreeable, conscientious versus unconscientious) intrinsically comprise a positive-negative contrast. Another positive aspect of MFT-IATs may be that, differently from other single category IATs (see above), the attention is drawn on the relevant category because the nonrelevant category includes various different attributes and can, therefore, not be used to simplify the task.

Coda

After the IAT procedure had been published (Greenwald et al., 1998), some psychologists feared that IATs were intended as a lie-detector revealing associations from the deep and inaccessible parts of personality that are more telling than what people tell about themselves. However, an important aspect of IATs is that they do not obscure the content of what is being assessed and that they allow participants to refuse cooperation. Although this raises a question about how “implicit” IATs are (cf. De Houwer & Moors, in press) this brings the important benefit that (different from procedures such as subliminal priming) IATs cannot be employed against the will of examinees. In fact, we are not aware of any publication exploring the validity of IATs that endorses a lie-detector view on this procedure.

IATs are beginning to be used as clinical research tools and may help to evaluate therapy needs and outcomes with regard to spontaneous behavior and automatic cognitive biases. For these purposes, IATs are not employed as self-sufficient procedures but as a useful adjunct to diagnosis via explicit self-reports. Moreover, it is clearly premature to consider IATs as tools for individual diagnosis in selection settings or as a basis for decisions that have important

personal consequences. The modest retest-reliability of IAT measures together with the unanswered questions concerning the explanation of IAT effects make evident that potential applications should be approached with care and scientific responsibility. Meanwhile, IATs are a fascinating research tool at the interface of social cognition and personality psychology that help to draw a more holistic picture of individual behavior and experience.

Footnotes

¹ We use the term anxiousness rather than anxiety to differentiate between the trait (anxiousness) and the state (anxiety) conceptualization of this construct.

References

- Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology, 91*, 652-661.
- 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. (2004). Measuring task-switching ability in the Implicit Association Test. *Experimental Psychology, 52*, 167-179.
- Banse, R. (1999). Automatic evaluation of self and significant others: Affective priming in close relationships. *Journal of Social and Personal Relationships, 16*, 803-821.
- Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie, 48*.
- 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.
- Blümke, M., & Friese, M. (2003). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology, 42*, 163-176.
- Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79*, 631-643.

- Boysen, G. A., Vogel, D. L., & Madon, S. (2006). A public versus private administration of the Implicit Association Test. *European Journal of Social Psychology, 36*, 845-856.
- Cvencek, D., & Greenwald, A. G. (2006). *Faking of the Implicit Association Test is statistically detectable*. Unpublished manuscript. University of Washington, Seattle.
- Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The Quad Model of implicit task performance. *Journal of Personality and Social Psychology, 89*, 469-487.
- Cronbach, L. J. (1957). The two disciplines of scientific Psychology. *American Psychologist, 12*, 671-684.
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12*, 163-170.
- Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for white Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology, 36*, 316-328.
- De Houwer, J. (2003a). A structural analysis of indirect measures of attitudes. In J. Musch, & K. C. Klauer (Ed.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 219-244). Mahwah, NJ: Lawrence Erlbaum.
- De Houwer, J. (2003b). The Extrinsic Affective Simon Task. *Experimental Psychology, 50*, 77-85.
- De Houwer, J., & De Bruycker, E. (in press). The Implicit Association Test outperforms the Extrinsic Affective Simon Task as an implicit measure of interindividual differences in attitudes. *British Journal of Social Psychology*.
- 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. (in press). How to define and examine the implicitness of implicit measures. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes*. New York: Guilford.
- 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., & Schmukle, S. C. (2003). Does social desirability moderate the relationship between implicit and explicit anxiety measures? *Personality and Individual Differences*, *35*, 1697-1706.
- Egloff, B., Schwerdtfeger, A., & Schmukle, S. C. (2005). Temporal Stability of the Implicit Association Test-Anxiety. *Journal of Personality Assessment*, *84*, 82-88.
- Ellwart, T., Becker, E. S., & Rinck, M. (2005). Activation and measurement of threat associations in fear of spiders: An application of the Extrinsic Affective Simon Task. *Journal of Behavior Therapy and Experimental Psychiatry*, *36*, 281-299.
- Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by redefining the category labels. *Journal of Experimental Social Psychology*, *40*, 357-365.
- Greenwald, A. G. (2005). *The Multifactor Trait IAT*. Unpublished manuscript. University of Washington, Seattle.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4-27.
- Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*, 1022-1038.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464-1480.

- Greenwald, A. G., & Nosek, B. A. (in press). Attitudinal dissociation: What does it mean? In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures*. Hillsdale, NJ: Erlbaum.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.
- Greenwald, A. G., Nosek, B. A., Banaji, M. R., & Klauer, K. C. (2005). Validity of the salience asymmetry interpretation of the IAT: Comment on Rothermund and Wentura (2004). *Journal of Experimental Psychology: General, 134*, 420-425.
- Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology, 90*, 1,20.
- 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.
- Huijding, J., & de Jong, P. J. (2005). A pictorial version of the Extrinsic Affective Simon Task. *Experimental Psychology, 52*, 289-295.
- Jordan, C. H., Spencer, S. J., Zanna, M. P., Hoshino-Browne, E., & Correll, J. (2003). Secure and defensive high self-esteem. *Journal of Personality and Social Psychology, 85*, 969-978.
- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 774-788.

- Karpinski, A., & Steinman, R. B. (2006). The Single Category Implicit Association Test as a Measure of Implicit Social Cognition. *Journal of Personality and Social Psychology*, *91*, 16-32.
- Kawakami, K., & Dovidio, J. F. (2001). The reliability of implicit stereotyping. *Personality and Social Psychology Bulletin*, *27*, 212-225.
- Kim, D. Y. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly*, *66*, 83-96.
- Klauer, K. C., & Mierke, J. (2005). Task-set inertia, attitude accessibility, and compatibility-order effects: New evidence for a task-set switching account of the Implicit Association Test effect. *Personality and Social Psychology Bulletin*, *37*, 208-217.
- Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (in press). Understanding and using the Implicit Association Test: IV. What we know (so far). In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies*. New York: Guilford Press.
- McConnell, A. R., & Liebold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, *37*, 435-442.
- McGregor, I., Nail, P. R., Marigold, D. C., & Kang, S.-J. (2005). Defensive pride and consensus: Strength in imaginary numbers. *Journal of Personality and Social Psychology*, *89*, 978-996.
- Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology*, *85*, 1180-1192.
- 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.

- Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*, 455-469.
- 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. (in press). Understanding the individual implicitly and explicitly. *International Journal of Psychology*.
- Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, *19*, 625-666.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, *31*, 166-180.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (in press). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes*. New York, NY: Psychology Press.
- Nosek, B. A., & Hansen, J. J. (2007). *The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation*. Unpublished manuscript. University of Virginia, Charlottesville.
- Nosek, B. A., & Smyth, F. L. (in press). A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology*.
- Nosek, B. A., & Sriram, N. (in press). Faulty assumptions: A comment on Blanton, Jaccard, Gonzales, and Christie (2006). *Journal of Experimental Social Psychology*.

- 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. (2005). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653-667.
- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*, 277-293.
- Penke, L., Eichstaedt, J., & Asendorpf, J. B. (2006). Single Attribute Implicit Association Tests (SA-IAT) for the assessment of unipolar constructs: The case of sociosexuality. *Experimental Psychology, 53*, 283-291.
- Perugini, M. (2005). Predictive models of implicit and explicit attitudes. *British Journal of Social Psychology, 44*, 29-45.
- Pinter, B., & Greenwald, A. G. (2005). Clarifying the role of the “other” category in the self-esteem IAT. *Experimental Psychology, 52*, 74–79.
- Poehlman, T. A., Uhlmann, E., Greenwald, A. G., & Banaji, M. R. (2007). *Understanding and using the Implicit Association Test III. Meta-analysis of predictive validity*. Manuscript submitted for publication.
- 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.
- Rudman, L. A., Greenwald, A. G., & McGhee, D. E. (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Journal of Personality and Social Psychology, 27*, 1164-1178.
- Schmukle, S. C. (2005). Unreliability of the dot probe task. *European Journal of Personality, 19*, 595-605.

- Schmukle, S. C., & Egloff, B. (2004). Does the Implicit Association test for assessing anxiety measure trait and state variance? *European Journal of Personality, 18*, 438-494.
- 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.
- Schmukle, S. C., & Egloff, B. (2006). Assessing anxiety with Extrinsic Simon Tasks. *Experimental Psychology, 53*, 149-160.
- 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 anger. *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., Wiesner, A., & Schütz, A. (in press). Self-esteem discrepancies and defensive reactions to social feedback. *International Journal of Psychology*.
- Steffens, M. C. (2004). Is the implicit association test immune to faking. *Experimental Psychology, 51*, 165-179.
- 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.
- Steffens, M. C., & Schulze-König, S. (2006). Predicting spontaneous Big Five behavior with Implicit Association Tests. *European Journal of Psychological Assessment, 22*, 13-20.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220-247.

- Teachman, B. A., & Woody, S. R. (2003). Automatic processing in spider phobia: Implicit fear associations over the course of treatment. *Journal of Abnormal Psychology, 112*, 100-109.
- 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.
- Wigboldus, D. H. J., Holland, R. W., & van Knippenberg, A. (2005). *Single target implicit associations*. Manuscript submitted for publication.

Table 1

Task Sequence and Stimuli of an Implicit Association Test to Measure the Implicit Personality Self-Concept of Anxiousness

Block	N of trials	Task	Response key assignment	
			Left key	Right key
1	20	Target discrimination	Me	Others
2	20	Attribute discrimination	Anxious	Confident
3	20	First block of first combined task	Me, anxious	Others, confident
4	40	Second block of first combined task	Me, anxious	Others, confident
5	40	Reversed target discrimination	Others	Me
6	20	First block of second combined task	Others, anxious	Me, confident
7	40	Second block of second combined task	Others, anxious	Me, confident
Categories	Target concept		Attribute concept	
	Me	Others	Anxious	Confident
Sample stimuli	Me	Others	Anxious	Confident
	I	They	Timid	Daring
	Self	Them	Insecure	Secure
	My	You	Worried	Unconcerned
	Own	Your	Cautious	Carefree

Note. Sample stimuli correspond to Schnabel et al. (2006b).

Table 2

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

Step	Procedure
1	Eliminate trials with latencies over 10,000 ms.
2	Exclude data from participants with more than 10% of trials showing latencies less than 300 ms.
3	Compute one “inclusive” standard deviation for all trials in Blocks 3 and 6 and likewise for Blocks 4 and 7.
4	Compute separate means for trials in each of the Blocks 3, 4, 6, and 7.
5	Compute two mean difference scores ($\text{Mean}_{\text{Block6}} - \text{Mean}_{\text{Block3}}$ and $\text{Mean}_{\text{Block7}} - \text{Mean}_{\text{Block4}}$).
6	Divide each difference score by its associated standard deviation of Step 3.
7	Resulting D measure represents the equal-weight average of the scores calculated in Step 6.

Note. This table is adapted from Table 3 in Lane, Banaji, Nosek, and Greenwald (in press).

An additional step is necessary if the IAT procedure does not prompt participants to correct errant responses (see text for details). Block numbers refer to Blocks described in Table 1.

SPSS and SAS syntax for this scoring algorithm are available at

http://faculty.washington.edu/agg/iat_materials.htm and www.briannosek.com.