Editorial

Advances and Challenges in the Indirect Measurement of Individual Differences at Age 10 of the Implicit Association Test

Wilhelm Hofmann[1] and Manfred Schmitt[2]

[1]University of Würzburg, Germany, [2]University of Koblenz-Landau, Germany

Ever since Sigmund Freud proposed that many core psychological processes cannot be self-observed because they occur outside of the conscious mind, psychologists have searched for methods that can be used to make diagnostic inferences about a person's dispositions without having to ask the person directly. On the one hand, the development of such indirect methods was driven by the assumption that persons may have only limited access to the underlying (automatic) processes that may be responsible for their thoughts, feelings, and actions (Nisbett & Wilson, 1977). On the other hand, the search for indirect measures was also motivated by the need to overcome a major weakness of direct (i.e., self-report) measures: their sensitivity to the distorting effects of the need for approval and other self-presentational concerns.

Despite many proposals of indirect measures, none received unanimous approval from the scientific community and none obtained the trust that researchers put in self-report measures despite their well-known shortcomings. The limited success of indirect measures can be accounted for by two major reasons: First, little agreement was achieved regarding the kinds of behavior that would be suitable indicators of unconscious processes and the inferences that could be drawn from them. For example, there is little consensus on what a specific comment on a TAT picture tells us about the personality profile of a person. Second, the reliability and validity of indirect measures tends to be disappointingly low when compared to direct measures.

With the fruitful exchange between cognitive and social psychology and the advance of computer-based assessment during the last 20 years, a new wave of mostly reaction-time-based indirect measures has been proposed (also referred to as *implicit* measures). Ten years ago, a general and flexible indirect assessment procedure appeared on the stage that seems to have considerable potential for overcoming the problems of previous indirect measures: the Implicit Association Test (IAT) as proposed by Greenwald, McGhee, and Schwartz (1998; see Schnabel, Asendorpf & Greenwald, 2008, for an up-to-date review). The IAT employs patterns in reaction time that people need for the classification of stimuli (words, pictures, symbols) according to their category membership. Classification is performed via simple motor responses such as pressing a key on a computer keyboard. The IAT is based on the assumption that classification can be performed faster if the classification of stimuli from *associated* concepts requires the same motor response. In the case of attitude measurement, the IAT presumes that attitudes are stored in memory as associations between an attitude category (nuclear energy) and an evaluative category (bad, evil). Accordingly, personality-IATs assume that traits are stored in memory as associations between the self and attribute concepts (talkative, intelligent, etc). The IAT is a very flexible procedure because it allows for the combination of a large number of concepts whose associations define the constructs to be measured.

The publication of the IAT had tremendous impact first on social cognition and attitude research from where it originated, and then soon after on several other research areas such as personality assessment, self-esteem research, special areas of clinical psychology, and consumer behavior. It instigated a renewed interest in indirect measures and promoted the development of other reaction-time-based indirect measures. It also directed attention toward the interplay of automatic versus controlled processes and helped to refine dual-process and dual-system conceptions of the mind. Finally, it prompted a new discussion of the consistency issue that had been a focus of debate in the attitude and personality literature three decades earlier.

As is the case with most new diagnostic paradigms, the IAT and other new indirect measures were not incorporated uncritically into the body of psychological assessment methods. Rather, many researchers submitted them to rigorous theoretical, methodological, and empirical analyses. These analyses identified a number of problems and open questions. Many of these ongoing issues will be addressed in the cut-

European Journal of Psychological Assessment 2008; Vol. 24(4):207–209
DOI 10.1027/1015-5759.24.4.207

ting-edge research presented in this special issue, which aims to advance our understanding of the IAT and related indirect measures and to contribute to their improvement.

1. *Reliability*. The IAT outperforms most other indirect procedures such as projective tests, priming measures, and the Emotional Stroop Test. However, two common strategies for estimating the reliability of an IAT do not converge. Internal consistency estimates are usually satisfactory (.80), whereas retest correlations are much lower (.55). The causes for this gap are not yet well understood although results of a few studies are suggestive (Gschwendner, Hofmann, & Schmitt, 2008a). The reliability of indirect measures is of particular focus in the model of response-interference task performance proposed by Gawronski, Deutsch, Le Bel and Peters (2008) and in the contribution by Rudolph, Schröder-Abé, Schütz, Gregg, and Sedikides (2008) in which reliability estimates based on internal consistency and retest correlations are compared across indirect self-esteem measures.

2. *Internal validity*. Several studies have shown that IATs and other indirect assessments may not measure only what they are intended to measure (e.g., the association strength between concepts); rather, these measures appear to be susceptible to method-specific sources of variance that are diagnostically irrelevant. For instance, a major problem of IATs, known as *recoding*, results from the possibility that respondents will classify IAT stimuli according to criteria other than the prescribed ones. As a consequence of such recoding strategies, the IAT's validity with regard to associations between target and attribute categories is compromised. This intriguing issue and related construct validity problems are discussed thoroughly in several papers (Gawronski et al., 2008; Nosek & Hansen, 2008; Schnabel et al., 2008; Teige-Mocigemba, Klauer, & Rothermund, 2008). The identification of validity problems has led researchers to develop procedural variants of the IAT intended to rule out these problems. One variant that specifically addresses the recoding problem is proposed by Teige-Mocigemba et al. (2008). These authors propose a Single Block IAT that makes recoding impossible. The first results reported in this issue are promising. Another procedural modification of the IAT was proposed by Olson and Fazio (2004). These authors argue that the construct validity of attitude IATs may be attenuated by individual differences in the knowledge of cultural stereotypes. Nosek and Hansen (2008) scrutinized whether the personalized IAT as proposed by Fazio and Olson is indeed superior to the classic IAT with regard to construct validity. Their answer is a clear "no" and they present abundant evidence suggesting that personalizing may increase a participant's likelihood of explicitly evaluating IAT target concepts (instead of classifying them according to task instructions).

3. *Convergent and discriminant validity*. One of the most pervasive findings in the literature that also becomes evident from virtually all of the empirical papers in this special issue pertains to the low convergence among indirect and direct measures. Correlations rarely exceed .30. The causes for this low convergence are not yet fully clear and have been a matter of controversial discussion. Some authors argue that indirect and direct procedures measure different mental representations that overlap only partially. This interpretation is consistent with dual-system or dual-process theories and implies that the IAT and other indirect measures possess divergent validity vis à vis direct measures. According to a second interpretation, the low convergence among indirect and direct measures indicates a validity problem. Given that the construct validity of established personality questionnaires has been proven, this interpretation implies that indirect measures lack construct validity. Note that this interpretation presumes that indirect and direct procedures measure a single construct. Unambiguous solutions to this issue are difficult to obtain because testing a theory (one vs. two distinct constructs) requires construct-valid measures. On the other hand, construct validation requires a theory. One paper discusses this problem in more detail than is possible here (Gschwendner, Hofmann, & Schmitt, 2008b). We believe that our understanding of the relationship between indirect and direct measures can be enhanced by taking into account theoretically relevant moderator variables (Gschwendner et al., 2008b; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Hofmann, Gschwendner, Nosek, & Schmitt, 2006; Nosek, 2005). Similar to previous consistency controversies, such an approach has already yielded the general conclusion that the direct-indirect relationship is highly conditional on a large range of personal, situational, and methodological boundary conditions. Thus, we need to abandon the view that there is a simple solution to the question of convergence or divergence, but rather develop sophisticated models based on the insights that these moderator effects offer regarding the dynamics and processes that underlie these measures.

4. *Predictive validity*. Another fruitful research strategy is to focus on actual behavior prediction as a criterion for evaluating the nature of indirect and direct measures. Finding that indirect measures have *incremental validity* over and above direct measures with regard to the prediction of behavior (e.g., Dewitte, De Houwer, & Buysse, 2008; Greenwald, Poehlman, Uhlmann, & Banaji, in press; Gschwendner et al., 2008b; Richetin & Perugini, 2008; Schmukle, Back, & Egloff, 2008) offers strong support against the contention that indirect measures lack construct validity. Furthermore, applying a moderator approach to the predictive validity of indirect measures (Gschwendner et al., 2008b; Hofmann, Gschwendner, Friese, Wiers, & Schmitt, in press; Richetin & Perugini, 2008) may provide us with new theoretical insights regarding *when* and *how* behavior is driven by indirectly measured representations.

5. *Convergence among different indirect measures*. Besides the moderator variables and the behavioral prediction approach, there is a third general strategy for obtaining a better understanding of the construct validity of indirect measures, that is, the degree of convergence among several indirect indicators. If indirect procedures really measure partially different constructs than do direct measures, several indirect measures of the same construct should converge. This strategy was pursued in one paper in this special issue. Rudolph et al. (2008) used six different indirect measures of self-es-

teem and correlated them with a self-esteem questionnaire. Despite the use of several strategies that improved the reliability of the indirect measures, their correlations with the self-esteem questionnaire did not differ reliably from zero. More importantly, and replicating results from a previous study (Bosson, Swann, & Pennebaker, 2000), the convergence among the indirect measures was rather low. Although this result is discouraging from the point of view of indirect self-esteem assessment, it is too early for an ultimate interpretation of the findings. However, it is very likely that a moderator approach focusing on procedural factors of indirect measurement may illuminate the reasons for the divergence/convergence of indirect measures and suggest potential ways to remedy this.

As editors we feel that the papers we assembled for this special issue satisfy three important selection criteria: First, we wanted to include papers that primarily address those issues that are most relevant for the assessment of individual differences: reliability and construct validity. Second, we wanted to include contributions that go beyond previous articles either methodologically or substantively. Third, we were looking for studies that both answer open questions and identify unresolved issues that need further research. We hope that our readers agree with us that the papers we have assembled tackle important issues and have the potential to stimulate further discussion and research. We are grateful to all authors for accepting our invitation and for sharing valuable insights from their research programs.

References

Bosson, J.K., Swann, W.B., Jr., & Pennebaker, J. (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.

Dewitte, M., De Houwer, J., & Buysse, A. (2008). On the Role of the Implicit Self-Concept in Adult Attachment. *European Journal of Psychological Assessment, 25,* 282–290.

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, 25,* 218–225.

Greenwald, A.G., McGhee, D.E., & Schwartz, J.L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

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

Gschwendner, T., Hofmann, W., & Schmitt, M. (2008a). 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.

Gschwendner, T., Hofmann, W., & Schmitt, M. (2008b). Convergent and Predictive Validity of Implicit and Explicit Anxiety Measures as a Function of Specificity Similarity and Content Similarity. *European Journal of Psychological Assessment, 25,* 254–262.

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., & Schmitt, M. (2006). What moderates implicit-explicit consistency? *European Review of Social Psychology, 16*, 335–390.

Hofmann, W., Gschwendner, T., Wiers, R., Friese, M., & Schmitt, M. (in press). Working memory capacity and self-regulatory behavior: Toward an individual differences perspective on behavior determination by automatic versus controlled processes. *Journal of Personality and Social Psychology.*

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 relationships between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134,* 565–584.

Nosek, B.A., & Hansen, J.J. (2008). Personalizing the Implicit Association Test increases explicit evaluation of target concepts. *European Journal of Psychological Assessment, 25,* 226–236.

Olson, M.A., & Fazio, R.H. (2004). Reducing the influence of extra-personal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653–667.

Richetin, J., & Perugini, M. (2008). When temporal contiguity matters: A moderator of the predictive validity of implicit measures. *European Journal of Psychological Assessment, 25,* 246–253.

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, 25,* 273–281.

Schmukle, S.C., Back, M.D., & Egloff, B. (2008). Validity of the Five-Factor Model for the Implicit Self-Concept of Personality. *European Journal of Psychological Assessment, 25,* 263–272.

Schnabel, K., Asendorpf, J.B., & Greenwald, A.G. (2008). Assessment of individual differences in implicit cognition: A review of IAT measures. *European Journal of Psychological Assessment, 25,* 210–217.

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, 25,* 237–245.

Wilhelm Hofmann

Department of Psychology
University of Würzburg
Röntgenring 10
D-97070 Würzburg
Germany
Tel. +49 931 312860
Fax +49 931 312812
E-mail hofmannw@psychologie.uni-wuerzburg.de