What Does the Implicit Association Test Measure? A Test of the Convergent and Discriminant Validity of Prejudice-Related IATs

Bertram Gawronski

University of Würzburg, Germany

Abstract. Drawing on recent criticism of the Implicit Association Test (IAT), the present study tested the convergent and discriminant validity of two prejudice-related IATs to corresponding explicit prejudice measures in a German student sample (N = 61). Confirming convergent validity, (a) an IAT designed to assess negative associations related to Turkish people was significantly related to the explicit endorsement of prejudiced beliefs about Turkish people, and (b) an IAT designed to assess negative associations related to East Asians was significantly related to explicit prejudice against East Asians. Moreover, confirming discriminant validity, (c) the Asian IAT was unrelated to the explicit endorsement of prejudiced beliefs about Turkish people, and (d) the Turkish IAT was unrelated to explicit prejudice against Asian people. These results further corroborate the assumption that the IAT is a valid method to assess the strength of evaluative associations in the domain of prejudice and stereotypes.

Key words: Implicit Association Test, prejudice, stereotypes, implicit cognition, validity

Recent research on prejudice and stereotypes has moved more and more away from investigating explicit prejudiced beliefs about members of stereotyped groups to exploring the automatic and unconscious cognitive processes that are suspected to be responsible for modern subtle forms of prejudice (Devine, 1989; Greenwald & Banaji, 1995). One of the central assumptions of these approaches is that negative or stereotypical associations relating to an out-group are activated automatically upon the encounter of a relevant stimulus (e.g., Devine, 1989). If a perceiver is not motivated or able to suppress these associations, his or her behavior will be directly affected by the activated evaluation or stereotype. If, however, the perceiver is motivated and able to control his or her automatic associations, the behavior will less likely be prejudiced or stereotypical (see also Fazio & Towles-Schwen, 1999).

Even though some models stress the generality of automatic attitude activation when people share the same cultural background (e.g., Bargh, Chaiken, Govender, & Pratto, 1992; Devine, 1989), more recent accounts assume that the automatic activation of attitudes and stereotypes differs as a function of the idiosyncratic strength of the respective associative links (e.g., Fazio, 1993; Greenwald & Banaji, 1995). Specifically, it is assumed that people differ with respect to the strength of negative or stereotypical associations relating to a given out-group, and hence exhibit different degrees of automatic attitude activation after encountering a relevant stimulus. Hence, one of the most important tasks in research on prejudice and stereotyping has been to develop measures

I would like to thank Rainer Banse for helpful comments on an earlier version of this article; Iain Glen for stylistic corrections; and Cornelia Eibach and Rebecca Wagner for their help in collecting the data.

DOI: 10.1027//1618-3169.49.3.171

that are capable of reliably assessing individual differences in the strength of negative or stereotypical associations (e.g., Fazio, Jackson, Dunton, & Williams, 1995; Greenwald, McGhee, & Schwartz, 1998; Nosek & Banaji, 2001; Wittenbrink, Judd, & Park, 1997).

The most widespread and also most controversial of these measures is probably the so-called Implicit Association Test (IAT) developed by Greenwald and his colleagues (Greenwald et al., 1998; see also the special issue of the *Zeitschrift für Experimentelle Psychologie, 48,* 2001). The IAT is based on a double discrimination task in which participants are asked to assign single stimuli (e.g., German and Turkish names) as fast as possible to a given pair of target categories (e.g., German vs. Turkish). Associative strength between two concepts is assessed by combining a given target dimension (e.g., German vs. Turkish) with an associated attribute dimension (e.g., positive vs. negative), both in an association-consistent and an association-inconsistent manner. The difference between the response latencies for association-consistent and association-inconsistent assignments is interpreted as an indicator for a participant's idiosyncratic associative strength between the two concepts. For example, evaluative associations towards Turkish people (as compared to Germans) may be assessed by combining the dimension "German vs. Turkish" and the dimension "negative vs. positive" in a prejudice consistent manner (i.e., Turkish – negative vs. German – positive) and a prejudice inconsistent manner (i.e., Turkish – positive vs. German – negative). The difference between the mean assignment latencies in these two tasks may then be interpreted as an index for the automatic activation of negative attitudes towards Turkish people as compared to Germans.

Notwithstanding its effective use in the prediction of nonverbal behavior (e.g., McConnell & Leibold, 2001), memory biases (e.g., Gawronski, Ehrenberg, Banse, Zukova, & Klauer, in press), known group membership (e.g., Banse, Seise, & Zerbes, 2001), sexual behavior (e.g., Marsh, Johnson, & Scott-Sheldon, 2001), or evaluative judgments (e.g., Neumann & Seibt, 2001), the IAT is still often criticized for being a theoretically obscure measure that possibly has nothing to do with associative strength. This criticism is mainly derived from recent studies on the cognitive processes that may be responsible for the performance in the IAT (e.g., Brendl, Markman, & Messner, 2001; Mierke & Klauer, 2001; Rothermund & Wentura, 2001).

Mierke and Klauer (2001), for example, presented evidence that IAT scores depend on differential task switch costs for the two combined tasks (Allport, Styles, & Hsieh, 1994; Rogers & Monsell, 1995). This, however, implies that spreading activation is insufficient to explain individual differences in the IAT. More seriously, obtained individual differences could be confounded with other variables related to task switching, such as intelligence or cognitive flexibility. Accordingly, the reported relations between IAT-scores and prejudice or stereotype relevant responses could also be due to a common relation to these variables, rather than to substantial differences in the strength of negative or stereotypical associations.

A similar conclusion may be drawn from Rothermund and Wentura's (2001) research on figure-ground asymmetries in the IAT. Rothermund and Wentura found that compatibility effects in the IAT can be switched as a function of which of the two target categories is considered as "figure" and which is regarded as "ground". As with the results of Mierke and Klauer (2001), this finding suggests that individual differences in the IAT might be confounded with other variables related to the development of task performance strategies such as intelligence or cognitive flexibility. Hence, the obtained relations between IAT scores and prejudice or stereotype relevant behavior could also be due to a common relation with these variables, rather than to individual differences in associative strength.

A third model for the emergence of IAT effects was presented by Brendl et al. (2001). According to Brendl et al., performance on the IAT can be explained on the basis of so-called random walk models (Townsend & Ashby, 1983). Basically, participants seem to consciously adapt their response criterion to the experienced difficulty of the combined tasks. Since such criterion shifts, however, can have multiple causes, an interpretation of IAT scores as individual differences in associative strength may be unwarranted. More seriously, the obtained relation between IAT scores and prejudice or stereotype relevant behavior could also be due to a common relation to variables affecting criterion shifts, rather than to individual differences in associative strength.

In sum, even though there is clear evidence for the *predictive* validity of the IAT (e.g., Gawronski et al., in press; McConnell & Leibold, 2001), studies on its *internal* validity seem to question an interpretation of these results in terms of individual differences in associative strength (e.g., Brendl et al., 2001; Mierke & Klauer, 2001; Rothermund & Wentura, 2001). In order to solve these interpretational ambiguities, the main goal of present research was to test the *convergent* and *discriminant* validity of the IAT, and thus to offer more insights into its general *construct* validity (Campbell & Fiske, 1959). For this purpose, we investigated the relation between two prejudice-related IATs and the explicit endorsement of prejudiced beliefs with respect to two different target categories in a multitrait-multimethod design.

Specifically, it was expected that prejudice-related IATs exhibit not only convergent validity, such that IAT scores are significantly related to the explicit endorsement of prejudiced beliefs, but also discriminant validity, such that IAT scores predict explicit prejudice only with respect to the same target category. If the IAT actually assesses individual differences in the strength of negative associations towards a given target category, IAT scores should necessarily exhibit both convergent and discriminant validity (Campbell & Fiske, 1959). If, however, the relation between IAT scores and prejudice-relevant behavior reflects a common relation to other variables such as intelligence or cognitive flexibility, the two IATs could, of course, exhibit convergent validity to explicit prejudice, but not discriminant validity.

Method

Participants

A total of 61 German students (39 female) drawn from a participants' volunteer pool took part in a study on personal beliefs about different ethnic groups. As an incentive for taking part, participants were paid 6 Euro. Experimental sessions were run in small groups with between 1 and 3 participants.

Explicit Prejudice Endorsement

In order to assess the explicit endorsement of prejudiced beliefs, participants were asked to complete a German version of Pettigrew and Meertens' (1995) Blatant Prejudice Scale related to Turkish and Asian people, respectively (Zick, 1997). Drawing on recent criticism of the subtle prejudice construct (e.g., Coenders, Scheepers, Sniderman, & Verberk, 2001; Gawronski, Petzold, & Banse, 2002), only items for blatant prejudice were included in the present study.[1] Blatant prejudice was assessed with 7 items using 5-point scales ranging from 1 (strongly disagree) to 5 (strongly agree).

Implicit Association Tests

Two IATs were developed to assess the strength of negative associations towards Turkish and Asian people. The IATs were run on three IBM-compatible Pentium III computers using the software DirectRT v2002 (Jarvis, 2001). In the first IAT, the target categories were Turkish and German; in the second IAT the target categories were Asian and German (see Table 1). Following Greenwald et al. (1998), both IATs consisted of five blocks, thus revealing a total of 10 blocks. In Block 1 (initial target-concept discrimination), 10 German and 10 Turkish names had to be assigned to the categories "German" or "Turkish", respectively (see Appendix). Participants were asked to press a left-hand key ("a") when a German name appeared on the screen, and a right-hand key ("5" of the number pad) in the case of a Turkish name. In Block 2 (attribute discrimination task), 10 positive and 10 negative nouns (Klauer & Musch, 1999; Schwibbe, Räder, Schwibbe, Borchardt, & Geiken-Pophanken, 1994; see Appendix) were presented and had to be classified according to the categories "positive" (left-hand key) and "negative" (right-hand key). In Block 3 (initial combined task), target and attribute discrimination trials were combined in a prejudice-consistent manner. Participants had to press the left-hand key when either a German name or a positive noun was presented, and the right-hand key when a Turkish name or a negative noun was presented. In Block 4 (reversed target-concept discrimination task), German and Turkish names had to be classified with a switch of the categorization keys. Block 5 (reversed combined task) again combined the two individual tasks, now in a prejudice-inconsistent manner. Participants had to press the left-hand key when either a Turkish name or a positive noun was presented, and the right-hand key when a German name or a negative noun was presented. The following five blocks were conceptually identical to Blocks 1 to 5, the only exception being the use of the category Asian rather than the category Turkish, and the presentation of Asian names in contrast to Turkish names (see Appendix). Each block started with a short instruction of the following task and a request to respond as fast as possible even if this would lead to errors. The discrimination tasks (Blocks 1, 2, 4, 6, 7, and 9) consisted of a total of 20 trials, respectively. The combined tasks (Blocks 3, 5, 8, and 10) each comprised 80 trials (40 names, 40 nouns), with targets and attributes being presented in an alternating order. Order

[1] Coenders et al. (2001) critized Pettigrew and Meertens' (1995) distinction between subtle and blantant prejudice for lacking empirical foundation. Specifically, Coenders et al. found evidence for a two-factor solution which substantially differend from that proposed by Pettigrew and Meertens (1995). The first factor was very broad and included items of both subtle and blatant prejudice. The second factor was rather small and mainly included items of the subtle prejudice scale. Moreover, whereas the first factor was well explained by a number of related variables, the second factor revealed only weak correlations to related constructs. Similar results were obtained by Gawronski et al. (2002), who found a high correlation of IAT scores to blatant prejudice, but only a low correlation to subtle prejudice.

Table 1. Task Sequences of Implicit Association Tests Used to Assess Negative Associations Towards Turkish People and Negative Associations Towards Asian People

Block	Task	Trials	Response Key Assignment	
			Left Key	Right Key
	German vs. Turkish			
1	Initial target discrimination	20	German	Turkish
2	Attribute discrimination	20	Positive	Negative
3	Initial combined task	80	Positive – German	Negative – Turkish
4	Reversed target discrimination	20	Turkish	German
5	Reversed combined task	80	Positive – Turkish	Negative – German
	German vs. Asian			
6	Initial target discrimination	20	German	Asian
7	Attribute discrimination	20	Positive	Negative
8	Initial combined task	80	Positive – German	Negative – Asian
9	Reversed target discrimination	20	Asian	German
10	Reversed combined task	80	Positive – Asian	Negative – German

of trials was randomized within each block and held constant for all participants. The response-stimulus interval following correct responses was 250 ms. Wrong responses were indicated with the word "FEHLER!" (German for "Error!"), which appeared for 1000 ms below the center of the screen.

Procedure

When participants arrived they were welcomed and informed that they were taking part in a study on personal beliefs about different ethnic groups. They were then seated in a cubicle in front of a computer. After a short instruction on the screen, participants were asked to respond to the items of Pettigrew and Meertens' (1995) Blatant Prejudice Scale. In the first part, items were related to prejudiced beliefs about Turkish people. A second block contained the same items related to Asian people. Items appeared on the screen one by one and were followed by the next item after participants had given their response. After completion of the explicit prejudice measures, participants were administered the two IATs. The first IAT was related to negative associations towards Turkish people, the second was related to negative associations towards Asian people. Finally, participants were debriefed and thanked for participation.

Results

Preliminary Analyses

Explicit Prejudice Endorsement. Items to assess blatant prejudice against Turkish and Asian people were merged into single indices by calculating mean values (Cronbach's $\alpha = .61$ for prejudice against Turkish people and .68 for prejudice against Asian people). Blatant prejudice against Turkish people ranged from 1.00 to 3.57 ($M = 1.92$, $SD = .53$); blatant prejudice against Asian people ranged from 1.00 to 3.43 ($M = 1.73$, $SD = .51$). The two indices were significantly correlated with $r = .55$ ($p < .001$).

Implicit Association Tests. Response latencies higher than 3000 ms were replaced by this value; latencies lower than 300 ms were recoded as missing values. Error trials were excluded from analyses. Individual IAT scores were calculated by first log-transforming response latencies and then subtracting the mean response times of the initial combined tasks from the mean latencies of the reversed combined tasks for each of the two IATs. Difference scores were calculated so that higher scores indicate more negative associations with respect to Asian or Turkish people. Raw latency scores for negative associations towards Turkish people ranged from −77 ms to 356 ms ($M = 107$, $SD = 93$), raw scores for negative associations towards Asian people ranged from −196 ms to 236 ms ($M = 55$, $SD = 79$). In order to estimate the reliability of the two IATs, the respective combined blocks were each divided into four sub-blocks of equal length (20 trials). These sub-blocks were then used to calculate four IAT scores for each attitude domain. Sub-scores for the Asian IAT revealed an internal consistency of .55; sub-scores for the Turkish IAT revealed an internal

consistency of .56 (Cronbach's α). The two IAT scores were significantly correlated with $r = .29$ ($p < .05$).

Convergent and Discriminant Validity

Table 2 presents the correlation matrix for the two IAT scores and the two indices of blatant prejudice. Confirming the convergent validity of the two IATs, (a) individual scores of the Turkish IAT were significantly related to blatant prejudice against Turkish people, and (b) individual scores of the Asian IAT were significantly related to blatant prejudice against Asian people. Furthermore, the discriminant validity of the two IATs was indicated by the lack of a significant correlation between (a) individual scores of the Asian IAT and blatant prejudice against Turkish people, and (b) individual scores of the Turkish IAT and blatant prejudice against Asian people. A Steiger's (1980) test of correlation differences revealed a significant difference between monotrait-heteromethod and heterotrait-heteromethod correlations for the Asian IAT, $z = 1.99$, $p < .05$, and a marginally significant difference for the Turkish IAT, $z = 1.37$, $p = .09$.

Table 2. Correlations Between Overall Scores of Negative Associations (Implicit) and Blatant Prejudice (Explicit) Related to Turkish People and Asian People

	1	2	3	4
1 Explicit Asian	(.68)	.55***	.32*	.21
2 Explicit Turkish		(.61)	.08	.37**
3 Implicit Asian			(.55)	.29*
4 Implicit Turkish				(.56)

Note.: Cronbach's α estimates of internal consistency are in parentheses; * $p < .05$, ** $p < .01$, *** $p < .001$

In order to offer a more stringent test of the predicted *double dissociation* implied by testing the convergent and discriminant validity, a structural equation analysis was conducted using LISREL 8.51 (Jöreskog & Sörbom, 2001). For this purpose, the two IATs were each divided into two halves by an odd-even split of the four subscores revealed in the estimation of internal consistencies (see preliminary analyses). In a similar vein, the two explicit measures were each divided into two halves by an odd-even split of the respective items. These observed variables were used to define four latent variables: implicit evaluations of Asian people (IAT-A), implicit evaluations of Turkish people (IAT-T), blatant prejudice against Asian people (BLAT-A), and blatant prejudice against Turkish people (BLAT-T) (see Table 3). The latent variables for implicit evaluations were defined as independent variables, and the latent variables for blatant prejudice were defined as dependent variables. Latent variables within the same method type (i.e., implicit vs. explicit) were allowed to correlate across the two attitude domains.

In order to test the strong version of the hypothesized double dissociation, a first model was tested that included paths from implicit evaluations to explicit prejudice only *within* a given attitude domain, but not *between* the two attitude domains (Figure 1). This most parsimonious model fitted the data very well, $\chi^2(16) = 19.58$, $p = .24$, RMSEA = .061, NNFI = .92, CFI = .95, and revealed highly positive path coefficients for both the Asian IAT ($\beta = .44$), and the Turkish IAT ($\beta = .52$). The correlation between the two IATs was estimated as $r = .44$; error covariance between the two explicit measures was estimated as .63.

In a second step, three less parsimonious models were defined that included either (a) a cross-path from implicit evaluations of Turkish people to blatant prejudice against Asian people, (b) a cross-path from implicit evaluations of Asian people to blatant prejudice against Turkish people, or (c) both possible cross-paths. Although these models were less restricted, they all fitted less well to the data than the strong version, implying a double dissociation, RMSEAs < .067, NNFIs < .90, CFIs < .95. Moreover, cross-path coefficients were close to zero for the Turkish IAT ($\beta = .06$ for the single cross-path, and .14 when both cross-paths were allowed), and negative for the Asian IAT ($\beta = -.22$ for the single cross path, and −.27 when both cross-paths were allowed).[2] These results corroborate the assumption that the two IATs actually tap different associative dimensions.

Discussion

The present results offer clear evidence for both convergent and discriminant validity of the IAT in the domain of prejudice-related associations. Specifically, the two IATs used in the present experiment were related to the explicit endorsement of prejudiced beliefs only when the out-group category in the IAT matched the target category in the explicit endorsement of prejudiced beliefs. Hence, it seems rather unlikely that the relation between IAT scores

[2] Since the Asian IAT was not correlated with blatant prejudice against Turkish people (see Table 2), this negative relation obtained in the structural equation model can be interpreted as a theoretically irrelevant suppressor effect driven by the significant correlation between the two IATs (Smith, Ager, & Williams, 1992).

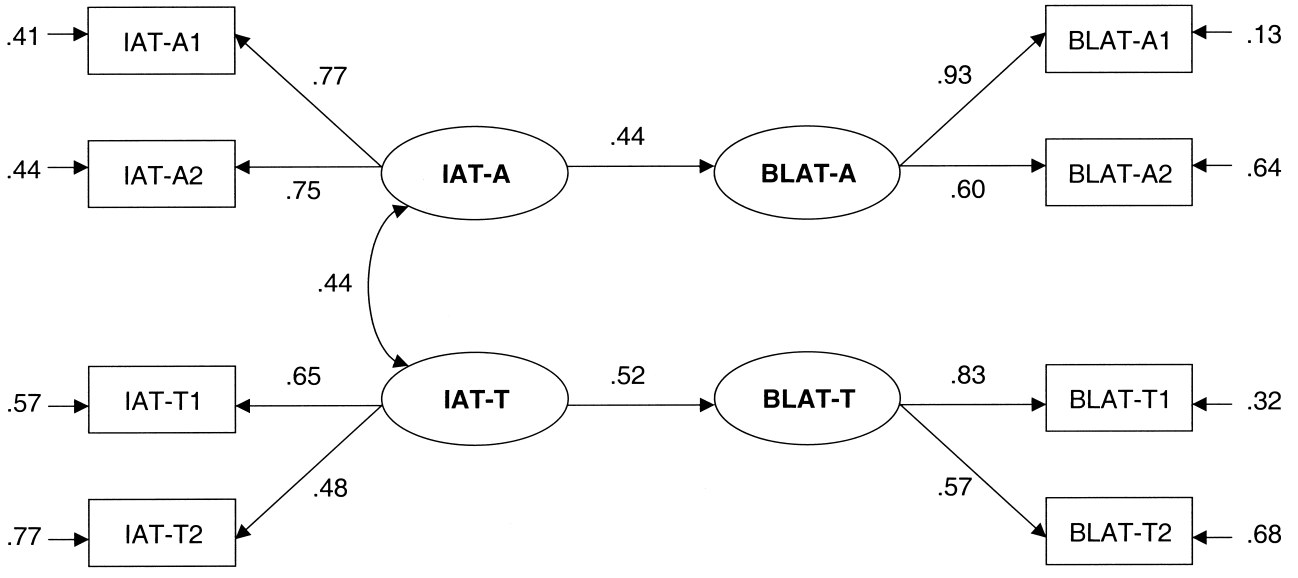

Figure 1. Structural equation model of blatant prejudice against Asian people (BLAT-A) and blatant prejudice against Turkish people (BLAT-T) predicted by negative associations towards Asian people (IAT-A) and negative associations towards Turkish people (IAT-T). 1 denotes the aggregation of items with an odd number of testing occasions; 2 denotes the aggregation of items with an even number of testing occasion. Standardized estimates by LISREL 8.51 (Jöreskog & Sörbom, 2001).

and prejudice or stereotype-related responses obtained in previous studies (e.g., Gawronski, et al., in press; McConnell & Leibold, 2001) can be attributed to a common relation to other variables such as intelligence or cognitive flexibility. Such a conclusion could have been drawn from the results of previous studies on the cognitive processes that are responsible for the performance in the IAT (e.g., Brendl et al., 2001; Mierke & Klauer, 2001; Rothermund & Wentura, 2001). The present findings clearly rule out this interpretational ambiguity. If the obtained relations between IAT scores and discriminatory responses actually stem from a common relation to other variables, such as intelligence or cognitive flexibility, the IAT may exhibit convergent validity, but not discriminant validity.

The present findings also extend previous results on the construct validity of the IAT. Cunningham, Preacher, and Banaji (2001), for example, found evidence for a substantial relation between the IAT and other implicit attitude measures, such as priming procedures and response window techniques. Moreover, these implicit measures revealed a significant relation to a corresponding explicit attitude measure. However, the measures used in the study of Cunningham et al. referred only to a single attitude dimension. Hence, it is not clear whether the IAT also exhibits discriminant validity when it comes to the assessment of different associative dimensions. The present findings demonstrate that this is actually the case.

Evaluative Associations?

Notwithstanding the present results, one may object that the obtained relations between IAT scores and the explicit endorsement of prejudiced beliefs could still be independent of evaluative associations. For example, being strongly immersed in a given culture could promote explicit prejudice against people who do not belong to this culture, and could independently lead to stronger figure-ground asymmetries resulting in higher IAT effects (see Rothermund & Wentura, 2001). In this case, the obtained correlations between the IAT and explicit measures may be spurious rather than systematically driven by associations. Such an interpretation, however, would have to explain how a strong cultural assimilation can affect figure-ground asymmetries in the IAT *independent* of associative strength. Since familiarity has already been ruled out as an alternative explanation for IAT effects (Dasgupta, McGhee, Greenwald, & Banaji, 2000; Ottaway, Hayden, & Oakes, 2001), and thus cannot account for culturally-based figure-ground asymmetries in the IAT, the most parsimonious explanation is still in terms of associations. Most importantly, differential task switch costs (Mierke & Klauer, 2001), figure-ground asymmetries (Rothermund & Wentura, 2001), or threshold adaptation (Brendl et al., 2001) could actually be driven by individual differences in associative strength. Specifically, it is possible that differential task switch costs in the IAT are higher for participants with strong

associations as compared to those with weak associations (see Mierke & Klauer, 2001). In a similar vein, for participants with strong associations, it may be the negative or stereotyped target category that "pops out" in terms of figure-ground asymmetries (see Rothermund & Wentura, 2001). Moreover, conscious threshold adaptation could be related to the experienced difficulty of the combined tasks (see Brendl et al., 2001), which, in turn, may vary as a function of associative strength. Drawing on this line of argument, these accounts may explain rather than question the relation between associative strength and IAT performance (De Houwer, 2001).

Another possibility that has not yet been discussed is that task switch costs (Mierke & Klauer, 2001), figure-ground asymmetries (Rothermund & Wentura, 2001), and conscious threshold adaptation (Brendl et al., 2001) are sources of systematic error variance in the IAT. In other words, the IAT may be affected by these processes even though they have nothing to do with associations. Note, however, that this does not imply that the IAT cannot be a valid measure of associative strength. Explicit prejudice measures, for example, are often affected by social desirability, but nevertheless reflect a large proportion of systematic variance caused by personal attitudes. In a similar vein, IAT scores could be affected by task switch costs, figure-ground symmetries, and threshold adaptation, but still reflect individual differences in associative strength. In any case, future research may help to further clarify the relation between associative strength and the cognitive processes that have been demonstrated to be relevant for the performance in the IAT.

Correlations to Explicit Measures

Another possible objection is that the obtained correlations between the two IATs and the explicit endorsement of prejudiced beliefs were only moderate as compared to, for example, the higher correlation between the two explicit prejudice measures. Moreover, different studies have reported quite different correlations between the IAT and related explicit measures. For instance, an ad hoc meta-analysis of recently published articles revealed a mean correlation of $r = .22$ ($SD = .17$), with correlations ranging from $-.18$ to .68 ($N = 180$). There are at least four explanations for these findings (see Banaji, 2001; Dovidio, Kawakami, & Beach, 2001 for more detailed discussions).

First of all, it is quite unlikely that explicit judgments generally reflect chronic associations in a one-to-one manner. Rather, explicit judgments often imply a number of processes that go far beyond just reporting one's associations, such as checks for representativeness (Strack, 1992), introspection about reasons (Wilson & Hodges, 1992), accessibility of exemplars (Smith & Zarate, 1992), or subjective experiences (Strack, 1992). Since these processes can differ as a function of the context and the judgmental domain, different contexts or different IATs can be expected to produce different correlations to corresponding explicit measures. Moreover, reliance on evaluative associations in explicit judgments about out-groups has been demonstrated to depend on a number of factors such as the motivation to control prejudiced reactions (e.g., Banse & Gawronski, 2001; Dunton & Fazio, 1997), awareness of the biasing influence of associations (e.g., Gawronski, Geschke, & Banse, 2002), the degree of cognitive elaboration (Florack, Scarabis, & Bless, 2001a), or perceived personal threat by the out group (Florack, Scarabis, & Bless, 2001b). Hence, differing correlations between prejudice-related IATs and explicit prejudice measures could also stem from the differing relevance of these variables for judgments about different out groups (e.g., Franco & Maass, 1999).

Second, it is important that implicit and explicit measures actually correspond to one another. Gawronski, Petzold, et al. (2002), for example, found a relatively high correlation between implicit evaluations and blatant prejudice, but only a weak relation between implicit evaluations and subtle prejudice (see Pettigrew & Meertens, 1995). Moreover, processing of prejudice-relevant information revealed identical effects on implicit evaluations and blatant prejudice, but had no impact on subtle prejudice. This result suggests that the dimensions of the blatant prejudice measure (i.e., perceived threat by the out group, overt rejection of the out group, and opposition to intimate contact with out group members) may be more directly related to implicit evaluations than the dimensions of the subtle prejudice measure (i.e., defense of traditional values, exaggeration of cultural differences, denial of positive emotions related to the out group). Most importantly, this seems to be true even though both measures were designed to assess explicit prejudice. Hence, when comparing correlations between implicit and explicit measures it is important to consider whether the two actually correspond to one another.

Third, correlations between implicit and explicit measures essentially depend on the reliability of the measures involved (Banaji, 2001; Dovidio et al., 2001). In the present case, for example, the internal consistencies of the two IATs were only moderate as compared to the rather high consistencies reported for some IATs used in previous studies (e.g., Banse et al., 2001; Gawronski et al., in press). Hence, it is possible that the reliability of the IAT differs as a function of its content. This would, in turn, explain why different IATs that ostensibly assess the same

Table 3. Means, Standard Deviations, and Correlation Matrix of Measured Variables

	IAT-A1	IAT-A2	IAT-T1	IAT-T2	BLAT-A1	BLAT-A2	BLAT-T1	BLAT-T2
IAT-A1	1.00							
IAT-A2	.58	1.00						
IAT-T1	.19	.25	1.00					
IAT-T2	.12	.29	.29	1.00				
BLAT-A1	.32	.22	.12	.16	1.00			
BLAT-A2	.21	.19	.16	.20	.55	1.00		
BLAT-T1	.12	.09	.28	.22	.58	.29	1.00	
BLAT-T2	−.02	.02	.28	.24	.29	.47	.47	1.00
M	.08	.06	.10	.15	1.91	1.50	2.04	1.76
SD	.11	.10	.13	.11	.61	.53	.61	.62

Note.: IAT-A = Implicit Association Test Asian, IAT-T = Implicit Association Test Turkish, BLAT-A = Blatant Prejudice against Asian people, BLAT-T = Blatant Prejudice against Turkish people; 1 denotes the aggregation of items with an odd number of testing occasions; 2 denotes the aggregation of items with an even number of testing occasions.

construct can exhibit different correlations to one and the same explicit measure. This importance of internal consistencies is also confirmed in the present study where the correlation between the two implicit measures increased from .29 to .44 after adjustment to measurement error in a structural equation model (see also Cunningham et al., 2001).

Finally, there are some methodological aspects related to the IAT that have the potential to attenuate correlations with corresponding explicit measures. On the one hand, the IAT seems to be affected by order effects of association-consistent and association-inconsistent combinations (Greenwald et al., 1998). Hence, it is often recommended to counterbalance the order of the two combined blocks if one wants to interpret the absolute value of the obtained difference scores. Even though this strategy may be adequate when the IAT is used as a dependent measure, it is actually not useful when the IAT is used as an independent measure. Specifically, due to the well-established order effects, counterbalancing can produce two different distributions that are not comparable to one another, i.e., one and the same difference score may represent a high score in one distribution and a low score in the other. Because collapsing such incomparable distributions can heavily reduce any correlation with other measures, counterbalancing is actually inadequate if one is interested in correlations with explicit measures. Similar problems arise when the order of the single trials are randomized for each participant anew, since this leads to a general confounding of individual differences in associative strength with individual order effects as a function of trial presentation.

Conclusion

In sum, the present findings corroborate the assumption that the IAT is a valid measure for evaluative associations in the domain of prejudice and stereotypes. In the present study, prejudice-related IATs exhibited not only convergent validity to corresponding explicit measures of prejudice endorsement, but also discriminant validity by revealing substantial relations only when the out-group category in the IAT was identical to that in the explicit measure. Hence, together with previous evidence for the predictive validity of the IAT (e.g., Gawronski et al., in press; McConnell & Leibold, 2001), this combination of convergent and discriminant validity further corroborates the construct validity of the IAT as a measure for individual differences in associative strength.

References

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421–452). Cambridge, MA: MIT Press.

Banaji, I. V. (2001). Implicit stereotypes and prejudice. In G. B. Moskowitz (Ed.), *Cognitive social psychology* (pp. 359–374). Mahwah, NJ: Erlbaum.

Banse, R., & Gawronski, B. (2001). *Motivation zur Vorurteilskontrolle: Skaleneigenschaften und Validierung* [Motivation to control prejudice: Scale properties and validation]. Manuscript submitted for publication.

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

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activa-

tion effect. *Journal of Personality and Social Psychology*, *62*, 893–912.
Brendl, C. M., Markman, A. B., & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*, 760–773.
Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.
Coenders, M., Scheepers, P., Sniderman, P. M., & Verberk, G. (2001). Blatant and subtle prejudice: Dimensions, determinants, and consequences. *European Journal of Social Psychology*, *31*, 281–297.
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. (2001, September). *Indirect measures of attitudes*. Paper presented at the 3rd Meeting of the European Social Cognition Network, Houffalize, Belgium.
Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.
Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 175–197). Malden, MA: Blackwell.
Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*, 316–326.
Fazio, R. H. (1993). Variability in the likelihood of automatic attitude activation. *Journal of Personality and Social Psychology*, *64*, 753–758.
Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–1027.
Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 97–116). New York: Guilford Press.
Florack, A., Scarabis, M., & Bless, H. (2001a). When do associations matter? The use of automatic associations toward ethnic groups in person judgments. *Journal of Experimental Social Psychology*, *37*, 518–524.
Florack, A., Scarabis, M., & Bless, H. (2001b). Der Einfluss wahrgenommener Bedrohung auf die Nutzung automatischer Assoziationen bei der Personenbeurteilung [The impact of perceived threat on the use of automatic associations in person judgments]. *Zeitschrift für Sozialpsychologie*, *32*, 249–259.
Franco, F. M., & Maass, A. (1999). Intentional control over prejudice: When the choice of measure matters. *European Journal of Social Psychology*, *29*, 469–477.
Gawronski, B., Ehrenberg, K., Banse, R., Zukova, J., & Klauer, K. C. (in press). It's in the mind of the beholder: The impact of stereotypic associations on category-based and individuating impression formation. *Journal of Experimental Social Psychology.*
Gawronski, B., Geschke, D., & Banse, R. (2002b). *Implicit bias in impression formation: Associations influence the construal of individuating information*. Manuscript submitted for publication.
Gawronski, B., Petzold, I., & Banse, R. (2002a). *Parallel context effects on implicit and explicit attitudes: Evidence for mutually independent processes*. Manuscript submitted for publication.
Greenwald, A. G., & Banaji, M. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.
Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.
Jarvis, W. B. G (2001). *DirectRT v2002*. New York: Empirisoft.
Jörsekog, K. G., & Sörbom, D. (2001). *LISREL 8.51*. Chicago, IL: Scientific Software International, Inc.
Klauer, K. C., & Musch, J. (1999). Eine Normierung unterschiedlicher Aspekte der evaluativen Bewertung von 92 Substantiven [A standardization of various aspects of the evaluation of 92 nouns]. *Zeitschrift für Sozialpsychologie*, *30*, 1–11.
Marsh, K. L., Johnson, B. T., & Scott-Sheldon, L. A. J (2001). Heart versus reason in condom use: Implicit versus explicit predictors of sexual behavior. *Zeitschrift für Experimentelle Psychologie*, *48*, 161–175.
McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, *37*, 435–442.
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.
Neumann, R., & Seibt, B. (2001). The structure of prejudice: Associative strength as a determinant of stereotype endorsement. *European Journal of Social Psychology*, *31*, 609–620.
Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, *19*, 625–664.
Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (2001). Implicit attitudes and racism: Effects of word familiarity and frequency of the Implicit Association Test. *Social Cognition*, *19*, 97–144.
Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in western Europe. *European Journal of Social Psychology*, *25*, 57–75.
Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207–231.
Rothermund, K., & Wentura, D. (2001). Figure ground asymmetries in the Implicit Association Test. *Zeitschrift für Experimentelle Psychologie*, *48*, 94–106.
Schwibbe, M., Räder, K., Schwibbe, G., Borchardt, M., & Geiken-Pophanken, G. (1994). Zum emotionalen Gehalt von Substantiven, Adjektiven und Verben [The emotional meaning of nouns, adjectives, and verbs]. In W. Hager, & M. Hasselhorn (Eds.), *Handbuch deutschsprachiger Wortnormen* (pp. 272–284). Göttingen, Germany: Hogrefe.
Smith, E. R., & Zarate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review*, *99*, 3–21.

Smith, R. L., Ager, J. W., & Williams, D. L. (1992). Suppressor variables in multiple regression/correlation. *Educational and Psychological Measurement*, *52*, 17–28.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251.

Strack, F. (1992). The different routes to social judgments: Experiential versus informational strategies. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 249–275). Hillsdale, NJ: Erlbaum.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.

Wilson, T. D., & Hodges, S. D. (1992). Attitudes as temporary constructions. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 37–65). Hillsdale, NJ: Erlbaum.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationships with questionnaire measures. *Journal of Personality and Social Psychology*, *72*, 262–274.

Zick, A. (1997). *Vorurteile und Rassismus. Eine sozialpsychologische Analyse* [Prejudice and racism: A social psychological analysis]. Münster, Germany: Waxmann.

Bertram Gawronski

University of Würzburg
Lehrstuhl für Psychologie II
Röntgenring 10
D-97070 Würzburg
Germany
Tel.: +49 931 312009
Fax: +49 931 312812
E-mail: gawronski@psychologie.uni-wuerzburg.de

Appendix A

Stimuli Used in the Implicit Association Tests

German Names	Turkish Names	Asian Names	Positive Nouns	Negative Nouns
Günther	Mehmet	Li	Heiterkeit	Ärger
Matthias	Kemal	Wang	Spaß	Elend
Harald	Ahmed	Zhao	Freundschaft	Hass
Stefan	Erkan	Zhang	Glück	Angst
Dieter	Özal	Peng	Freude	Unglück
Eberhard	Murat	Chang	Gesundheit	Verrat
Wolfgang	Abdullah	Wu	Liebe	Streit
Volker	Ali	Quian	Paradies	Pest
Michael	Mohammed	Feng	Begeisterung	Krankheit
Konrad	Mustafa	Jiang	Entspannung	Panik

