

Clarifying the Role of the "Other" Category in the Self-Esteem IAT

Brad Pinter¹ and Anthony G. Greenwald²

¹The Pennsylvania State University, Altoona College, ²University of Washington, Altoona, PA, USA

Abstract. A. Karpinski (2004) recently criticized Implicit Association Test (IAT) measures of self-esteem, arguing that their measurements of self-associations are compromised by their contrasting *self* with a putatively extremely negative second category, the *nonspecific other*. The present data show, to the contrary, that the *nonspecific other* category in the self-esteem IAT is near neutral in valence. Validity of the self-esteem IAT is most appropriately assessed by examining its correlations with conceptually related measures. That has been done in several previous studies that are reviewed here. The *nonspecific other* category is only one of several choices for representing the concept of *other* in self-esteem IATs. Choice of the appropriate *other* category to contrast with *self* in self-esteem IATs should be guided by the needs of the research question being addressed.

Keywords: implicit self-esteem, Implicit Association Test (IAT), measurement

Karpinski (2004) recently proposed that the self-esteem IAT does not provide a valid measure of implicit self-esteem. His critique consisted of two related points. His first point was that the "IAT is not simply a measure of self-esteem; it is a joint measure of self- and other-esteem" (p. 23). Karpinski was referring to the inclusion of both the concept *self* and the nonspecific concept *other* in self-esteem IATs. Karpinski described the dependence of IAT scores on both target concepts as "restrictive" (p. 23) and a "general problem with all IAT research" (Footnote 2, p. 33). Karpinski's second point was that "the finding that most people have very positive implicit self-esteem when measured with the esteem-IAT may not be due to high implicit self-esteem, but the comparison of the self to a very negative other" (p. 30). After indicating our perspective on these two points of Karpinski's critique, we report an experiment that provides data bearing on the second point.

The Relative Character of the IAT

According to Karpinski (2004), a central problem with the IAT is its dependence on relative associations of concepts with attributes. In fact, the relative nature of an IAT measure would be problematic only if one wished (inappropriately) to interpret an IAT effect as

if it depended on associations with just one of the two target concepts. Karpinski attributes such an error of interpretation generally to users of the self-esteem IAT in suggesting that "the standard interpretation of the esteem-IAT is that it measures the associations one has with the self" (p. 5). To the contrary, the standard interpretation of any IAT measure involves relative strengths of associations of the two contrasted concept categories with the two contrasted attribute categories (Greenwald, McGhee, & Schwartz, 1998).

The Contrasting Target Concept in the Self-Esteem IAT

Because of the relative nature of the IAT, it is important that the target concept chosen to contrast with *self* be selected in a way that is conceptually appropriate for the research question. There is a long tradition in research on the self of using the *nonspecific other* category as a suitable contrast. Not surprisingly, therefore, the nonspecific other category has been used often in self-esteem IATs. However, several alternative contrast categories have been used in IATs that assess associations with self. In the research of Greenwald and Farnham (2000) and Greenwald et al. (2002) both *generic* self-other formats (self and other both represented by pronouns) and *idiographic* self-other for-

mates (self and other represented by first name, last name, hometown, etc.) were used and found to produce similar findings. Yamaguchi and colleagues (e.g., Chen, Yamaguchi, & Greenwald, 2003) developed two additional self-esteem IATs for a cross-cultural study of self-esteem that contrasted self with *ingroup* and self with *best friend*. We see the flexibility in the choice of contrasting target categories not as a weakness, but as a useful property of the IAT.

In an IAT, the two contrasting target concepts may have equal variability of their association strengths with attributes used in the IAT, such that both categories contribute equally to individual difference variance in resulting IAT scores. However, the more general case is that the two target concepts will differ in the variability of strengths of their associations with the IAT's attribute concepts. Karpinski (2004) erroneously assumed that only the former case is possible in arguing that the "esteem-IAT reveals as much about one's other-associations as it does about one's self-association" (p. 23). In principle, a self-esteem IAT could assess variation in association strengths involving just the self category if there were no variation across persons in strengths of associations with the target category chosen to contrast with self. Of course, establishing the degree to which variability in a self-esteem IAT is due to associations involving self vs. associations involving the contrasted other is a matter to be established in research, rather than something to be asserted a priori.

On the basis of results from two experiments, Karpinski (2004) questioned the appropriateness of the nonspecific other category for use in a self-esteem IAT. Both of Karpinski's experiments involved comparisons of a self-other IAT (i.e., using pronouns to represent both self and other) with variants in which the nonspecific other was replaced with known positive categories, *close friend* (Study 1) and *Santa Claus* (Study 2), or with a known negative category, *Adolf Hitler* (Study 2). Karpinski reported that self-other IAT measures were more positive (i.e., showing greater positive valence associated with self) than either self-close friend or self-Santa Claus IAT effects. He argued that "this result provides clear evidence that the esteem-IAT is influenced by the content of the other" (p. 27) and "that changing the other in the esteem-IAT also changes the nature of the self" (p. 30). These results are neither surprising nor informative regarding the appropriateness of measuring implicit self-esteem with the IAT. As has already been established, the IAT is a measure of relative associations of concepts and attributes. The fact that changing one of the concepts or attributes in an IAT affects the IAT effect is expected under a wide variety of assumptions

about how the IAT works, and has no direct bearing on the validity of IAT measures.

Karpinski (2004) also reported that self-other IAT effects were not statistically different from self-Hitler IAT effects. From this he concluded that "although this [IAT] method cannot reveal the *content* of the unspecified other, it can be inferred that the *valence* of the other ... was not different from the valence of Hitler" (p. 29). This conclusion is problematic partly because it is an inappropriate acceptance of a null hypothesis and also because of some of the methods used in collecting and analyzing data (these are noted in presenting our own experiment, below). However, the greatest concern about this conclusion is that it was not based on any attempt to directly contrast the valence of the Hitler and nonspecific other categories. Assessing that contrast was the primary aim of the present experiment.

Method

Subjects completed three IAT measures that assessed relative valence associations, respectively, of self and other, self and Hitler, and other and Hitler. The first two of these IATs were included for comparison with the data of Karpinski (2004); the third provided a direct assessment of the valence associations of the nonspecific other category relative to Hitler.

Subjects

Thirty-nine undergraduates (23 female, 16 male) at the University of Washington participated in exchange for optional course credit. Complete data were available for all subjects.

Procedure

Subjects were seated in separate rooms containing a computer. After providing consent and demographic information, subjects learned that they would be classifying words representing a variety of categories, including self, other, *pleasant*, *unpleasant*, and Hitler.

Subjects then completed three seven-block IAT measures (Greenwald et al., 1998) that were designed to measure relative valence associations of three concept pairs – self and other, self and Hitler, and other and Hitler. All IATs used pleasant and unpleasant attributes, represented by five words each for the pleasant (*cheer*, *happy*, *health*, *laughter*, and *peace*)

and unpleasant (*death, filth, jail, murder, and sickness*) categories. Five words were also used to represent each of the target concepts – self (*I, me, my, mine, and self*), other (*others, they, them, their, and theirs*), and Hitler (*Adolf, Germany, Gestapo, Auschwitz, and Nazi*).¹

The number of trials in each IAT block was identical for the three IATs (24 in Blocks 1 and 2 (practice on concept and attribute classifications), 36 in Block 5 (reversal of concept classification), 24 in Blocks 3 and 6 (practice combined-task blocks), and 40 in Blocks 4 and 7 (test combined-task blocks). The additional trials in Block 5 were included to reduce potential effects of the order of the two combined-task conditions as described by Nosek, Greenwald, and Banaji (in press). Concept and attribute trials were alternated on combined-task blocks and subjects were required to correct errors in order to continue, with latencies recorded to provision of the correct response. Subjects completed the three IATs in one of two orders: self-other, other-Hitler, self-Hitler or the reverse. The order of the combined-task blocks within each IAT was counterbalanced such that each IAT was done in both possible orders equally often, across subjects.

Results

Data Reduction

A recently developed improved IAT scoring procedure was used (see Greenwald, Nosek, & Banaji, 2003). This procedure uses data from practice and test combined-task blocks as well as individual-subject latency variability from those blocks to compute IAT scores that have some characteristics of individual-subject effect size measures. Use of the new *D* measure notwithstanding, the present conclusions would be unaltered if results were reported using the log-transform versions of the IAT measure as introduced by Greenwald et al. (1998) and used by Karpinski (2004).

The present research did not use Karpinski's (2004) analysis procedure of making ad hoc adjustments of IAT measures to correct for order effects (see Karpinski's Footnotes 3 and 5) or of conducting analyses on IAT trial subsets (see Karpinski's Tables

3 and 6). Order effects are more suitably dealt with by counterbalancing order of IATs in the research design, as in the present experiment. The trial-subset analysis strategy used by Karpinski was based on the inappropriate assumption that subsets of IAT trials can provide information about strengths of associations of individual target concepts with attribute categories. A recent analysis of large Web IAT data sets by Nosek et al. (in press, Study 1) indicated that this is an invalid strategy – the IAT remains a relative measure in its various subsets of trials.

IAT Magnitudes

Table 1 displays means and other descriptive information for the three IATs.² For the self-other IAT, subjects responded more rapidly on blocks in which category pairings grouped self with pleasant and other with unpleasant (than in blocks with the complementary pairings), $M = 0.56$ ($SD = .32$), $F(1, 37) = 119.47$, $p = 10^{-13}$, Cohen's $d = 1.77$. This pattern of results represents the typical IAT marker of positive implicit self-esteem (Greenwald & Farnham, 2000).

For the self-Hitler IAT, subjects were substantially faster on blocks that grouped self with pleasant and Hitler with unpleasant, $M = 0.71$ ($SD = .39$), $F(1, 37) = 190.82$, $p = 10^{-16}$, $d = 1.83$. There was an additional effect on the magnitude of the self-Hitler IAT effect resulting from the order of administration, $F(1, 37) = 19.66$, $p = 10^{-5}$. Table 1 shows that self-Hitler IAT effects were larger (.93 vs. .48) when that measure was administered first than when it was administered last. This effect of IAT order on the self-Hitler IAT is, of course, also apparent on the contrast between the self-other and self-Hitler IATs, $F(1, 37) = 15.32$, $p = 10^{-4}$. The difference between self-other and self-Hitler IAT scores was significantly different from zero only when subjects completed the self-Hitler IAT first, $F(1, 37) = 19.85$, $p = 10^{-4}$, $d = 1.17$ (otherwise, $F[1, 37] = 1.26$, $p = .27$, $d = .23$). This latter finding replicates a result reported by Karpinski (2004), whereas the former result, suggesting that nonspecific other and Hitler have very different valences, is inconsistent with his findings. Because Karpinski did not counterbalance (i.e., the self-Hitler IAT was always

¹ Karpinski's (2004) items to represent Hitler were *Hitler, Nazi, Germany, WWII, and Jews*. Including Jews as an IAT item to represent the category *Hitler* seems as inappropriate as using *Whites* as an IAT item to represent the category *African American*. Karpinski's choice of *WWII* to represent Hitler also seemed inappropriate because of the many associations of WWII to concepts with valence very different from Hitler. Consequently, these two items were replaced with the presumably more suitable items, *Auschwitz* and *Gestapo*.

² Effect sizes computed using milliseconds (not shown), although descriptively consistent, were uniformly smaller than those for the *D* measure, in agreement with the previous observations of Greenwald et al. (2003).

Table 1. Descriptive statistics for self-other, self-Hitler, and other-Hitler IATs.

IAT	N	Mean	SD	Reliabilities	Correlations	
					Self-Hitler	Other-Hitler
<i>Overall</i>						
Self-other IAT	39	.56	.29	.42**	.24	-.17
Other-Hitler IAT	39	.37	.35	.24	-.01	
Self-Hitler IAT	39	.71	.39	.56**		
<i>Self-Hitler 1st</i>						
Self-other IAT	20	.56	.29	.45*	.44 ⁺	.12
Other-Hitler IAT	20	.31	.34	.12	.09	
Self-Hitler IAT	20	.93	.31	.33		
<i>Self-other 1st</i>						
Self-other IAT	19	.57	.35	.49*	.22	-.42 ⁺
Other-Hitler IAT	19	.44	.36	.33	.16	
Self-Hitler IAT	19	.48	.33	.48*		

Note. Table values use *D* metric (Greenwald, Nosek, & Banaji, 2003). Positive means indicate that responses were faster on trials in which self + pleasant (for the self-other and self-Hitler IATs) and other + pleasant (for the other-Hitler IAT) shared a response key than on trials with complementary pairings. Reliabilities were computed by correlating measures computed separately from practice and test blocks of each IAT.

⁺ $p < .10$; * $p < .05$; ** $p < .01$.

completed last) it is unknown whether his results were due to incomplete counterbalancing.

If nonspecific other and Hitler have difference valences, why was that difference not apparent when the self-Hitler IAT came last in sequence? One perspective is that these results are unsurprising in light of previous demonstrations of the effect of recent experience on IAT effects (Greenwald et al., 2003; Greenwald & Nosek, 2001) and of the reduction of IAT effects with prior IAT experience (Greenwald et al., 2003; Nosek, Banaji, & Greenwald, 2002). Therefore, Karpinski's non-finding of a difference between the self-other and self-Hitler IATs may be an artifact of his procedure. What these explanations do not address, however, is the absence of a similar order effect on the self-other IAT. In light of potential ambiguity resulting from the order effect, the most definitive evidence comes from a between-subject comparison that only includes IATs that were completed first. That analysis revealed a statistically significant difference between the self-Hitler and self-other IATs, $F(1, 37) = 11.36$, $p = .002$, $d = 1.08$.

Although the present findings suggest that the valence of Hitler is more negative than the valence of

other, more decisive evidence for that conclusion comes from the direct comparison of valences of nonspecific other and Hitler provided by the other-Hitler IAT. Table 1 shows that subjects were faster on blocks that paired other with pleasant and Hitler with unpleasant, $M = 0.37$ ($SD = .35$), $F(1, 37) = 44.82$, $p = 10^{-8}$, $d = 1.06$. This result conflicts with Karpinski's (2004) conclusion that the valence of the nonspecific other concept is as negative as that of Hitler. To the contrary, the other-Hitler IAT shows that the valence of nonspecific other was more positive than that of Hitler by an amount corresponding to more than the conventional designation of a large effect size.³

IAT Correlations and Reliabilities

Table 1 shows descriptive information for the three IAT measures. Half-test reliabilities for the self-other and self-Hitler IATs were slightly lower than those reported in prior studies but, given small sample sizes, are not significantly divergent. However, the other-Hitler IAT showed quite low reliability, suggesting that there was relatively little systematic individual

³ One can question whether this is an appropriate conclusion given that the meaning of nonspecific other (or any category used in an IAT) may change depending upon the identity of the category with which it is contrasted. As an example, when *man* is contrasted with *boy* in an IAT, the property of being an adult is prominent in the meaning of man as a category label, and the features that men share with boys (i.e., maleness) are less important. In contrast, when man is contrasted with *woman*, the features shared by men and boys (i.e., maleness) should be quite prominent in the meaning of man. For the same reason, when other is contrasted with self its meaning may not be the same as when it is contrasted with Hitler, although this difference seems more difficult to specify than in the man example. Note that Karpinski's (2004) conclusion regarding the valence of other rests on the tenuous assumption that the meaning of self remains invariant across the self-other and self-Hitler IATs.

variability in this IAT. Three of the largest intercorrelations among measures involved the self-other and self-Hitler IATs, and the largest of these correlations (.44), was approximately an upper limit as constrained by the measures' reliabilities. Similarly, one of the largest correlations in Karpinski's (2004) data involved the self-other and self-Hitler IATs ($r = .38$; $N = 48$; Table 7 on p. 30). Karpinski suggested that this correlation implies that the valence of other and Hitler are similar. Despite the inherent difficulties in interpreting correlations between two relative measures, we believe it is equally possible (indeed more plausible) that the measures are related because both measures involve contrasts with the focal category self. Thus, on this perspective the observed correlations between the self-other and self-Hitler IATs reflect the fact that both measures capture variability in valence associations with self. With respect to the other category, five out of the six correlations involving the other-Hitler IAT showed only slight deviations from zero, suggesting that valences of either or both of other and Hitler categories vary little across subjects (with other presumably being uniformly near neutral and Hitler being uniformly negative).

Additional Evidence on Valence of the Nonspecific Other

In an additional data collection, 41 subjects from the same population as the main experiment provided data for an IAT that included the two valence attributes (pleasant and unpleasant) along with the contrast between nonspecific other (represented with the same 5 words as previously) and *middle* (represented by *center*, *halfway*, *average*, *middle*, and *midway*).

The category middle was selected to contrast with other because middle was previously found to be useful as a relatively inert contrast category in IATs, allowing IAT scores to be determined largely by associations involving the category with which it is contrasted (McGhee, 2001). This usefulness of middle depends on the plausible (although empirically unverified) assumption that middle has relatively few associations of its own with the attribute categories often used in IATs. It is possible that neither target concept has much inter-individual variability in association strengths with valence attributes. Although this reasoning undermines use of the middle-other IAT as an individual difference measure, the sample mean for this IAT should nevertheless provide a satisfactory

comparison between average strengths of associations with valence for middle and nonspecific other.

Administration of the middle-other IAT was procedurally similar to that of the IATs of the main experiment. Results for the middle-other IAT showed that the valence of middle was slightly more positive than the valence of other. For the *D* measure, $M = .05$ ($SD = .34$), $F(1, 40) = .84$, $p = .37$, $d = .14$. These results indicate that the valence of nonspecific other is statistically indistinguishable from that of middle. In turn, this suggests that the valence of nonspecific other may be approximately neutral. However, in the absence of sure knowledge of the valence of middle, this assertion rests mainly on general understanding that middle is a concept that is relatively neutral in valence.⁴

Discussion

Karpinski (2004) implied that the validity of the self-esteem IAT depends on the valence of the concept of nonspecific other. Although, as shown here, Karpinski's conclusion about the valence of nonspecific other was incorrect, establishment of that valence is actually more relevant for interpreting numerical values of IAT measures than it is for appraising the validity of the self-esteem IAT. Rather, evidence in the form of correlations with conceptually relevant other measures is needed to appraise the validity of any IAT. Correlational evidence for the IAT's validity was reviewed by Greenwald and Nosek (2001). Poehlman, Uhlmann, Greenwald, and Banaji (2004) included many more recent studies in a meta-analysis that demonstrates the IAT's predictive validity, especially in domains in which impression management is likely to distort self-report measures. Karpinski reported low correlations of his self-esteem IAT with explicit self-esteem measures, which was consistent with previous findings (Bosson, Swann, & Pennebaker, 2000; Greenwald & Farnham, 2000). However, Karpinski presented no correlational evidence bearing on convergent or predictive validity of the self-esteem IAT. Fortunately, previous investigations have presented such correlations. For example, Greenwald and Farnham (2000) showed that the IAT was reliably associated with reactions to experimentally induced successes and failures. Bosson et al. (2000) found that the self-esteem IAT was significantly correlated with indicators of esteem, certainty, and competence in

⁴ The reliability of the other-middle IAT was a poor $r = .30$, which is not unexpected if neither target category showed much variability in relation to the valence attributes. This is yet another reason to suspect the valence of other is approximately neutral.

self-descriptive essays. At the same time, these observed correlations have not been large.

Additional evidence of predictive validity of self-esteem IATs has been slow to accumulate, at least partly because of the difficulty of identifying suitable behavioral or other criteria for validating self-esteem measures. This difficulty of establishing predictive validity is true also for long-used measures of explicit self-esteem. A recent review of literature on explicit self-esteem revealed variable (and typically modest) correlations with behaviors across a variety of domains (Baumeister, Campbell, Krueger, & Vohs, 2003).

In conclusion, current evidence indicates that non-specific other remains a useful category for self-esteem and self-concept IATs. At the same time, evidence for the value of alternative categories to use as a contrast with self in IAT measures is accumulating. Depending on the specifics of research questions, these alternatives – which included best friend, in-group, and various specific others – will sometimes be preferable.

Acknowledgement

This research was supported by grants from National Institute of Mental Health, MH-41328, MH-01533, and MH-57672. The authors are grateful to Laurie A. Rudman and Brian A. Nosek for comments on an earlier draft.

References

- Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest*, 4, 1–44.
- Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and elephant revisited? *Journal of Personality and Social Psychology*, 79, 631–643.
- Chen, D. H., Yamaguchi, S., & Greenwald, A. G. (2003, June). *Cross-cultural comparisons of implicit and explicit self-esteem*. Paper presented at the Northwest Cognition and Memory Conference, Seattle, WA.
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S., (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, 109, 3–25.
- Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, 79, 1022–1038.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464–1480.
- Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift fuer Experimentelle Psychologie*, 48, 85–93.
- Greenwald, A. G., Nosek, B., & 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.
- Karpinski, A. (2004). Measuring self-esteem using the Implicit Association Test: The role of the other. *Personality and Social Psychology Bulletin*, 30, 22–34.
- McGhee, D. E. (2001). Unpublished data. University of Washington.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting intergroup implicit attitudes and beliefs from a demonstration website. *Group Dynamics*, 6, 101–115.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (in press). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*.
- Poehlman, T. A., Uhlmann, E., Greenwald, A. G., & Banaji, M. R. (2004). Understanding and using the Implicit Association Test: III. A meta-analysis of predictive validity. Unpublished manuscript, Yale University.

Received June 1, 2004

Revision received July 29, 2004

Accepted July 30, 2004

Address for correspondence

Brad Pinter
Penn State Altoona
3000 Ivyside Park
Altoona, PA 16601
USA
Tel. +1 814 949 5507
Fax +1 814 949 5011
E-Mail tbp1@psu.edu