The Associations in our Heads Belong to Us: Searching for Attitudes and Knowledge in Implicit Evaluation

Brian A. Nosek and Jeffrey J. Hansen

University of Virginia

Authors' note: This research was supported by a grant from the National Institute of Mental Health (R01 MH68447-01). Thanks to Mahzarin Banaji, Jerry Clore, Tony Greenwald, Fred Smyth, Bethany Teachman, Tim Wilson and Mark Zanna for helpful comments. Correspondence concerning this article should be addressed to Brian Nosek, Department of Psychology, University of Virginia, 102 Gilmer Hall, P.O. Box 400400, Charlottesville, VA 22911. Electronic mail may be sent to nosek@virginia.edu. Related information can be found at http://briannosek.com.

Abstract

Explicitly, humans can easily distinguish their own attitudes from evaluations possessed by others. Implicitly, the viability of a distinction between personal (attitude) and extra-personal (evaluative knowledge) is less clear. We investigated relations between self-reported attitudes, evaluative knowledge and the Implicit Association Test (IAT). Further, we tested IAT procedural modifications that are purported to reduce its sensitivity to extra-personal knowledge. In seven studies, we observed that (a) the attitude-IAT relationship varies from moderate to strong, (b) the knowledge-IAT relationship varies from little to none, and (c) the proposed IAT procedural changes do not diminish a knowledge confound but they do foster a task-recoding confound. Ownership of mental associations is established by presence in mind and influence on thinking, feeling and doing, not by explicit avowal or disavowal of the associations. The distinction between *personal* and *extra-personal* will need some conceptual refinement if it is to provide a useful taxonomy for implicit evaluation.

For intentional thoughts and actions there is little debate over their source. Intended acts are products of the self via psychological mechanisms like goals, attitudes, and beliefs. But what of thoughts and actions that are unintended? In one sense, it is a tautology to say that, whether intended or unintended, the activities of a brain and mind belong to the individual taking residence in that brain. In another sense, the question highlights potential confusion over the proper attribution when intentions and actions are dissociated. A committed egalitarian may find herself prejudging a Hispanic job applicant as unqualified. She may be distressed by those thoughts as inconsistent with her 'self', her ideology, and her honest attitudes toward Hispanics. What then is the source of her prejudgment – to whom does it belong?

Automaticity has taken hold in psychology with a broad range of mental life now understood to proceed without the encumbrances of consciousness, intention, and control (Bargh, 1996; Wegner & Bargh, 1998). The notion of automaticity affirms that the 'owner' of actions is the individual, but simultaneously points out that the instigation of action need not be an act of conscious will that is experienced as coming from the self (Wegner, 2002). The consciously-experienced self may be just another observer of the daily activities of the body it inhabits, perhaps having only slightly more privileged access than a self existing in another body (Wilson, 2002). This provides some basis for comprehending the protest "I did it, but it wasn't me." Thoughts and actions may come from my body and brain, but not always with the stamp of self-approval.

Devine (1989), for example, showed that egalitarians and non-egalitarians alike automatically associated Blacks with negative stereotypes. The discrepancies between explicitly intended and automatically assessed evaluations have spurred confusion about how (or whom) to attribute the automatic associations. Should automatic evaluations be considered a reflection of the person, even if they are consciously rejected? Or, perhaps, they provide little insight about the person and instead showcase the cultural context in which the person is embedded.

In this paper, we examine the extent to which automatic associations, as measured by the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), assess meaningful individual differences in evaluations versus knowledge of others' evaluations. Also, we evaluate IAT modifications

that were proposed to eliminate a contaminating influence of extra-personal knowledge. We demonstrate that (a) the IAT corresponds with self-reported attitudes and not with evaluative knowledge, (b) the proposed IAT procedural modifications alter assessment of attitudes but do not alter the IAT's relationship with evaluative knowledge, and (c) the procedural modifications increase the influence of a confounding factor that disrupts implicit attitude assessment.

Theory and Evidence Concerning the Presence of Attitudes and Knowledge in Automatic Evaluation

In her seminal demonstration of automatic racial biases, Devine (1989) understood the automatic associations assessed by her evaluative priming procedure to reflect knowledge as a 'culturally-shared stereotype', whereas explicit measures of racial animus were reflective of personal attitudes. In other words, self-report assessed individual differences in racial attitudes, and the priming procedure assessed cultural knowledge that was not reflective of individual differences.

Claiming that a measure administered to individuals reflects information about the culture and not the individual suggests that a cultural construct is represented at a subordinate level of analysis (the individuals living in that culture). This implies there is no meaningful variability in automatic evaluations across individuals because the variability would show that, by definition, the variation in automatic bias reflects an individual difference, not a cultural constant. As Fazio, Jackson, Dunton, and Williams (1995) explained: "If, as Devine suggested, the shared cultural stereotype is activated in the presence of a minority group, one would expect little meaningful variation in the pattern of facilitation across participants. On the other hand, if it is one's personal evaluation that is activated in the presence of a minority group member, the variation across participants would be more substantial and predictive of race-relevant behaviors" (p. 1015). Indeed, Fazio et al. demonstrated that individual variation in automatic racial evaluations was associated with rated friendliness and interest during an interaction with a Black experimenter. This observation is critical in pointing out that the meaningful variation across individuals indicates that the automatic evaluation reflects something about persons rather than groups or cultures.

In retrospect, data from Devine (1989) showed variability in perceptions of stereotypes. In the first study, participants reported the cultural stereotype about African-Americans. Far from consensus, not a single characteristic of the stereotype was generated by all participants. In fact, most qualities (e.g., low intelligence, uneducated, sexually perverse) were mentioned by 20% - 50% of the respondents indicating substantial variability in the perception of cultural stereotypes. Real people aside, even social psychologists exhibit variability in perceptions of racial animus in present day American culture, with some arguing that prejudice is still widespread (e.g., Sears, in press), and others arguing that it is vastly overestimated by social scientists (e.g., Arkes & Tetlock, in press). From a cultural psychology perspective, these data come as little surprise because culture is not a singular construct recorded invariantly across minds, individuals have unique, personal experiences of their cultural context (see Fiske, Kitayama, Markus, & Nisbett, 1998).

The fact that there is variation in perceptions of cultural evaluations, suggests that the mental store of evaluative knowledge includes evaluations that are believed (my attitudes) and evaluations that are known but not believed (evaluative knowledge). Explicitly, it is clear that these two sources of information are distinguishable. It is easy for someone to report that he hates green eggs and ham and will not eat them, and simultaneously report that Sam likes them and should eat them himself. With awareness and control, one can opt to use his own evaluations to guide judgment and behavior, and choose *not* to use knowledge about others' evaluations.

The fact that evaluative knowledge can be selectively discounted means that people may possess evaluative knowledge about the world that is self-irrelevant in that it does not guide responses to the target of evaluation. So, even though evaluative knowledge is variable across individuals, this may provide a basis for distinguishing personal evaluations (attitudes) from extra-personal ones (knowledge), the former being the evaluations that direct individual action, and the latter being knowledge that is not used for individual responses to attitude objects (Karpinski & Hilton, 2001; Olson & Fazio, 2004).

When individuals are allowed to report their evaluation, the act of reporting is itself a personal statement. But, when evaluation is assessed implicitly, the self-relevance of the evaluation is less clear.

Perhaps implicit measures are susceptible to evaluative information that does not actually guide individual behavior and, instead, just exists in mind as knowledge about the world.

Attitudes and knowledge in the Implicit Association Test. In examining the Implicit Association Test, Karpinski and Hilton (2001) concluded that it measured knowledge, not attitudes. This conclusion was based on their demonstration that the IAT correlated weakly with self-reported attitudes toward flowers relative to insects, a domain thought not to elicit strong self-presentational concern, and a null relationship between Candy Bar-Apple IAT evaluations and the participants' subsequent choice of a candy bar or apple. These results suggested that the IAT is not sensitive to individual attitudes, so Karpinski and Hilton concluded that the task reflects evaluative knowledge that is irrelevant to feeling and acting and therefore non-attitudinal.

A conclusion that the IAT reflects *only* extra-personal evaluative knowledge does not survive against the weight of evidence. A recent meta-analysis of 81 studies demonstrated that the IAT is moderately correlated with self-reported attitudes and (r = .24; Hofmann, Gawronski, Gschwendner, & Le, 2004). Further, in a study of 57 different attitudinal domains, Nosek (2005) reported that the strength of the correlation between the IAT and self-reported attitudes varied from near zero for some attitude domains (e.g., Thin People-Fat People) to approximately .70 in other domains (e.g., Pro Choice-Pro Life). Also, in a meta-analysis of 61 studies investigating the predictive validity of the IAT, Poehlman, Uhlmann, Greenwald, and Banaji (2004) found that the IAT predicted a wide variety of judgments and behaviors (r = .27).

These data demonstrate that the IAT is an individual difference measure and is associated with individual-level thoughts, feelings, and actions. However, they do not rule out the possibility that evaluative knowledge *also* influences IAT scores. Olson and Fazio (2004) proposed that extra-personal associations included any evaluative knowledge that is irrelevant to the personal attitude. They did not make a distinction between personal and extra-personal on the basis of endorsement of the evaluation as Karpinski and Hilton did, but rather on the basis of what information is activated from memory in response to the attitude object. Further, Olson and Fazio proposed that the original form of the IAT is

"contaminated by associations that do not contribute to one's evaluation of an attitude object and thus do not become activated when one encounters the object but that are nevertheless available in memory" (p. 653).

As evidence for the contaminating influence of extra-personal associations on the IAT, Olson and Fazio (2004) pointed out that a greater proportion of individuals show an automatic preference for White relative to Black on the IAT than do on evaluative priming, another measure of automatic evaluation. This was interpreted as evidence that the IAT was overestimating the magnitude of personal racial bias and must be influenced by evaluative knowledge, which they presumed to be more negative against Blacks than personal evaluations.

Rather than measuring evaluative knowledge directly and showing that it predicted IAT scores, Olson and Fazio (2004) proposed procedural modifications to the IAT to reduce the assumed influence of extra-personal associations and then demonstrated that the original and modified IATs elicited distinct effects. For one, the modified IAT appeared to elicit weaker racial bias estimates making it more similar to results observed with evaluative priming, a measure presumed to be uninfluenced by extra-personal associations. A second line of evidence used self-reported attitudes as a criterion variable. Olson and Fazio predicted that if the procedural modifications were effective at reducing the confounding influence of extra-personal knowledge, then the modified IAT should relate more strongly to the attitude criterion than would the original IAT. In two studies measuring political (Gore-Bush) and food (Candy Bar-Apple) attitudes, the modified IAT showed stronger relations with self-reported attitudes than did the original IAT.

Searching for Attitudes and Knowledge in Implicit Evaluation

We agree with Olson and Fazio (2004) that the distinction between personal and extra-personal should not be made on the basis of personal endorsement. At the same time, we are not convinced that the current evidence supports an interpretation that the IAT is influenced by extra-personal knowledge or that the IAT modifications alter that influence. Most critically, the relationship between evaluative knowledge and the IAT has not been demonstrated directly. Without measuring or manipulating

evaluative knowledge, it is difficult to determine if it is related to the observed difference between original and modified IATs.

In seven studies across a variety of content domains, we (a) examined the IATs relationship with evaluative knowledge, (b) tested whether proposed IAT modifications altered the IAT-knowledge relationship, and (c) tested our hypothesis that observed differences between the original and modified IATs are due to a confounding factor introduced (not removed) by the procedural modifications.

Our studies generated four distinct lines of evidence regarding the relations among attitudes, knowledge, and the IAT:

(1) *Unique attitudinal variation in IAT versions* – In Studies 1-3, we tested whether the IAT modifications altered attitude assessment instead of (or in addition to) removing extra-personal variation. Both modified and original IATs showed evidence of measuring unique attitudinal variation, suggesting that the procedural modifications were doing something other than reducing contaminating variation.

(2) *Relations between attitudes, knowledge, and the IAT* – In Studies 2-7, we measured explicit attitudes and evaluative knowledge directly and examined their relations with the IAT. We found substantial evidence linking attitudes with the IAT, and little evidence to support the presumed link between knowledge and the IAT.

(3) *Moderation of knowledge relation with IAT procedural modifications* – In Studies 2-7 we found no evidence to support the prediction that the relationship between evaluative knowledge and the IAT would be diminished by the proposed procedural modifications, casting further doubt on the hypothesized mechanism for the difference between original and modified IATs.

(4) *Identifying a confounding factor in IAT modifications* – In Studies 4-7 we found evidence for our hypothesis that, instead of removing a confound of evaluative knowledge, the procedural modifications actually introduced a confound of task-recoding in which the target concepts were more likely to be explicitly evaluated.

Study 1 – Political Attitudes

Olson and Fazio (2004) observed stronger correlations between self-reported attitudes and the modified IAT compared to the original IAT. They interpreted this as evidence that the modified IAT removed extra-personal contaminating variance in the original IAT, thus bolstering its relation with self-reported attitudes. We agree that such a difference in correlations may be a necessary condition for showing the reduction of contaminating variance, but it is not sufficient to reveal the identity of the contaminating variance, nor does it require a conclusion that removal of contaminating variance is the operative cause.

One alternative is that the procedural modifications change attitude assessment and have little to do with altering the influence of evaluative knowledge or any other contaminating variable. If the procedural modifications just remove contaminating variance from the IAT, then any attitudinal variation in the original IAT should be redundant with that measured by the modified IAT. If, on the other hand, the procedural modifications alter attitude measurement, then both IATs may contain attitudinal variation that is non-redundant with the other IAT. Following Olson and Fazio's (2004) use of explicit attitudes as a criterion variable to index attitudinal variation in the IAT, we tested whether original and modified IATs would predict explicit attitudes after removing the common variance between them.

Method

Participants

Ninety-six University of Virginia undergraduates (ages 17-22) participated in the study for partial course credit. Prior to analysis, three participants were removed because one did not complete all of the measures and two were not fluent English speakers. Initial analysis removed two participants for high error rates (>20%) in the standard evaluative priming task, and nine participants who did not follow task instructions in performance of the modified IAT. Of the remaining 82 participants, 58 were female and 24 were male; 65 were White, 8 were Asian, 5 were Black, 2 were Hispanic, and 2 were a different ethnicity.

Materials

Stimulus items. Ten exemplars representing the evaluative categories directly replicated the normative items used by Olson and Fazio (2004). Four stimulus items (two head shots, last name, first and last name) represented the categories 'George Bush' and 'Al Gore' for all of the implicit measures. The items were the same as ones used for the 2000 Presidential Election task at the Implicit Association Test Web site (https://implicit.harvard.edu/; Nosek, Banaji, & Greenwald, 2002a).

Implicit Association Test. The procedure for the IAT (both original and modified) replicated the versions used by Olson and Fazio (2004) described in their fourth study. The original and modified IATs were the same except for two procedural differences. One difference involves changing the evaluative category labels from ones thought to emphasize normative judgments in the original IAT (Good/Bad, Pleasant/Unpleasant) to ones that emphasize idiosyncratic judgments in the modified IAT (I Like/I Dislike). The other change eliminated error feedback for incorrect responses for the modified IAT with the assumption that such feedback reinforces making normative judgments rather than idiosyncratic evaluations of each evaluative exemplar. All other procedural details were identical and are described below.

Participants completed seven blocks of response trials for each of the two IATs. First, participants sorted evaluative words for 20 trials into categories (Pleasant/Unpleasant for the original IAT; I like/I dislike for the modified IAT) using two response keys on a standard keyboard. Second, using the same response keys participants sorted faces and words associated with Bush and Gore for 20 trials into categories (Bush/Gore). Third, participants sorted items for all four categories (Bush, Gore, Pleasant [I like], Unpleasant [I dislike]) for 20 trials using the two response keys. One key was used to categorize Gore and Pleasant [I like] items; the other key was used to categorize Bush and Unpleasant [I dislike] items. Fourth, the same key mapping was repeated for 40 more trials. Fifth, like the 2nd block, participants sorted Bush and Gore items again for 20 trials except that the response mapping was reversed (i.e., if Gore items were categorized with the left key before, they were now categorized with the right key). Sixth, again participants sorted items from all four categories for 20 trials except that the response mappings for the category exemplars (Bush/Gore) were opposite of the 3rd and 4th blocks. So, in this

example, Gore and Unpleasant [I dislike] were sorted with one key and Bush and Pleasant [I like] were sorted with the other. And, seventh, participants repeated the sorting conditions in the 6th block for 40 more trials.

In blocks with four categories, trials alternated between presenting category (Gore, Bush) and attribute (Pleasant [I like], Unpleasant [I dislike]) items. Also, reminder labels appeared at the top of the screen for all blocks reminding participants of the current sorting task. Further, to emphasize the distinction between the category and attribute dimensions, "Gore/Bush" labels and items appeared in black, and "I like/I dislike" labels and items appeared in green. For the original IAT only, categorization errors were identified with a red 'X' below the stimulus item and participants had to correct the response before continuing to the next trial. An interstimulus delay of 150 milliseconds separated each trial. Finally, the order of the category mapping conditions (Gore with Unpleasant [I dislike] before or after Gore with Pleasant [I like]) was counterbalanced between-subjects.

IAT analysis followed recommendations of Greenwald, Nosek, and Banaji (2003). Data were divided into three parcels, each comparing 20 trials from the two response conditions, which served as manifest indicators of a latent IAT factor. The IAT scores were coded such that positive values indicated liking for Gore relative to Bush.

Explicit measures. Explicit evaluations of Al Gore and George Bush were measured with five different explicit measures (adapted from Olson and Fazio, 2004) and are described in the appendix. All explicit measures were coded such that positive values indicated a preference for Gore over Bush.

Procedure

Participants completed four implicit measures and a set of explicit measures assessing evaluations of Al Gore and George Bush. Presentation of measures were counterbalanced as follows: half of the participants completed the explicit measures before the implicit measures and the other half completed them in the opposite order; within implicit measures, half of the participants completed the modified tasks before the original tasks and the other half completed them in the opposite order; within task type, half of the participants completed the IAT before priming and the other half completed them the in opposite

order; and within IATs, participants performed the category pairings in a between-subjects counterbalanced order. Results from the evaluative priming tasks are not discussed here (see supplements available at http://briannosek.com).

Results and Discussion

In this first study, we sought to replicate Olson and Fazio's (2004) demonstration that the modified IAT corresponded more strongly with self-reported attitudes than did the original IAT. Also, we tested whether both the original and modified IATs contained unique attitudinal variation. This finding would suggest that the procedural modifications are indeed altering attitude measurement, but they are not just removing confounding variation, if at all.

As an initial demonstration we calculated correlation coefficients between latent factors representing the original IAT, modified IAT, and explicit attitudes. The IAT factors were related to one another (r = .55) and both were strongly related to explicit attitudes (attitude-original IAT r = .61; attitude-modified IAT r = .72). These correlations are consistent with Olson and Fazio's (2004) demonstration that the IAT-attitude correlation was somewhat stronger for the modified IAT (z-bar = 1.53, p = .13; Steiger, 1980).[1]

To test whether the two IAT versions each contained unique attitudinal variation, we regressed explicit attitudes on the original and modified IATs simultaneously. The resulting estimates reflected the IAT-attitude relation for each IAT version after removing the shared variance. As predicted, both the modified (β = .55, $p < .001$) and original (β = .31, p = .004) IATs had significant regression coefficients predicting explicit attitudes suggesting that each measure contained unique attitudinal variation.[2]

The change in correlations between IAT versions has been interpreted as indicating that the modified IAT removed the influence of extra-personal associations from the original IAT. Results from Study 1 suggest that while the modified IAT was more strongly related to self-reported political attitudes, this may be because it changes attitude assessment, and does not necessarily reduce or eliminate contaminating variance. Even so, it is still possible that the modifications change attitude assessment and

remove contaminating variance simultaneously. There is yet no compelling alternative explanation for why the IAT-attitude correlation should be stronger following procedural modifications.

Study 2 – Racial Attitudes

Study 1 demonstrated that the two versions of the IAT each captured unique attitudinal variation, but left unexamined the question of whether extra-personal knowledge is influencing the IAT and is affected by the procedural modifications. In Study 2, we sought to replicate the demonstration from Study 1 in a different content domain (attitudes toward Blacks relative to Whites) and to begin testing whether extra-personal associations in the form of evaluative knowledge influence the IAT and are influenced by the procedural modifications.

Hypothesis Testing Through Comparisons of Nested Structural Equation Models

Comparing results of nested structural equation models allows us systematically to test specific hypotheses about the structure of relations between variables (McArdle, Johnson, Hishinuma, Miyamoto, & Andrade, 2001). For example, we can test whether the original and modified IATs differ in their relation to evalautive knowledge by comparing the fits of two differently constrained models of the interrelationships among knowledge, explicit attitudes, and the IAT. The models are identical except that one constrains the functional relation between knowledge and each of the IAT versions to be equal, while the other model allows those relations to vary freely for the modified and original IATs. If the latter model demonstrates a significant improvement in overall model fit, then we can conclude that relations between the two versions of the IAT and knowledge are significantly different. This enables a direct test of the conceptual question without isolating that comparison from the interrelationships with other variables (as would be the case with a significance test of the difference in correlation coefficients).

All models share a common structure in which explicit attitudes and evaluative knowledge are factors predicting the IAT. In Studies 2 and 3, original and modified IATs were collected within-subjects and are represented as two endogenous variables in a single model. In Studies 4-6, original and modified IATs were measured between-subjects and so we conducted two-group (original or modified) structural equation modeling with a single endogenous IAT variable (McArdle & Hamagami, 1996).

Assessment of model fit was based on the *root-mean-square error of approximation* index (RMSEA or ε_a;Browne & Cudeck, 1993; Steiger & Lind, 1980). This index weighs absolute fit, which declines whenever a parameter is removed from the model, against model complexity, such that the benefits of parsimony are considered along with fit (Steiger, 2000). Models fitting with $\varepsilon_a < .05$ are usually considered "close" fits, .05 to .08 as "fair" fits, .08 to .10 as "mediocre," and above .10 as "poor" (MacCallum, Browne, & Sugawara, 1996). Change in model fit from one proposed model to another was assessed by $\Delta\chi^2/\Delta df$ and by the 95% confidence interval (CI) of ε_a generated by this change.

Method

Participants

One hundred forty-nine University of Virginia undergraduates (ages 17-22) participated in the study for partial course credit. Prior to analysis, one participant was removed because of a computer malfunction. Initial analysis removed six participants for high error rates (>20%) in at least one of the response latency tasks. Of the remaining 142 participants, 99 were female and 43 were male; 109 were White, 15 were Asian, 11 were Black, 8 were Hispanic, 8 were a different ethnicity, and 1 did not report ethnicity.

Materials

Stimulus items. Four Black faces and four White faces taken from the IAT demonstration website (https://implicit.harvard.edu/; Nosek, et al., 2002a) served as exemplars for the 'Black American' and 'White American' categories in the IAT, and as the primes for the evaluative priming task. Again, the evaluative priming results are not discussed here but do appear in supplementary materials.

Implicit Association Test. The procedure for the Implicit Association Test (both original and modified) replicated the version used by Olson and Fazio (2004; Study 4) and in Study 1 of the present paper. The single change to the procedure was to use the categories and stimulus items representing 'Black American' and 'White American' rather than 'Gore' and 'Bush'. The IAT score was coded such that positive values indicated liking for White Americans relative to Black Americans.

Explicit measures. Explicit attitudes were assessed with self-reported feelings of warmth and liking of Black and White Americans. Evaluative knowledge was assessed with ratings of the historical favorability of society toward Black and White Americans, the favorability of cultural portrayals of the racial groups, and American society's warmth for the racial groups (*r*s among knowledge measures ranged from .36 to .52). All items and descriptive statistics are presented in the appendix. Participants also completed the 13-item Racial Arguments Scale (Saucier & Miller, 2003), but it was unrelated to all of the other measures and is not considered further. Items were coded such that positive values indicated a stronger positivity toward White Americans relative to Black Americans.

Procedure

Presentation of measures followed the same format described in Study 1.

Results

Unique Attitudinal Variation in Modified and Original IATs

We first sought to replicate the observation from Study 1 that original and modified IATs each contained unique attitudinal variation. A structural model in which explicit attitude was regressed on original and modified IATs showed that the modified IAT contained significant unique attitudinal variation ($\beta = .27$, $p = .02$), whereas the original IAT retained a non-significant amount of unique attitudinal variation in common with self-report ($\beta = .20$, $p = .08$). This is a weak replication of Study 1 in the domain of racial attitudes. This may be because the relationship between self-reported racial attitudes and IAT scores tends to be relatively weak (Nosek et al., 2002a), there is less available shared variance for multiple measures to explain. In Study 3, we sought to replicate this effect one more time in the domain of food attitudes.

Relations between the IAT and explicit attitudes and evaluative knowledge

The primary hypotheses concerning the relations between attitudes, knowledge and the IAT were tested through a comparison of nested structural equation models. We began with an unconstrained model in which all parameters were freely estimated and proceeded to add constraints in accord with specific hypotheses, testing whether the constraints resulted in significant losses in overall model fit. The

final model for Study 2 is presented in Figure 1, and results of the sequence of nested models are presented in the top panel of Table 1.

Question 1: Is knowledge differentially predictive of original and modified IATs (comparison of Models 0 and 1)? We hypothesized that procedural modifications to the IAT would not affect the relationship between the IAT and evaluative knowledge. If that is true, then constraining equal the paths between knowledge and the two IAT versions should have little impact on overall model fit. Adding such a constraint (Model 1) resulted in little change in χ^2 ($\Delta\chi^2(1) = 0.2$, $p = .65$), and overall model fit is actually slightly improved ($\varepsilon_a = .063$ from $\varepsilon_a = .065$). The 95% confidence interval around the *RMSEA of the change* in fit (95% CI ε_a of Δ) between these models includes .05, a further indicator that the model fits are very close to one another.

Question 2: Is evaluative knowledge related to the IAT at all (comparison of Models 1 and 2)? We also predicted that evaluative knowledge would be unrelated to IAT scores, and tested this question by comparing Model 1 with a model in which the relationship between evaluative knowledge and the IAT was fixed to zero (Model 2). If constraining this relationship to zero resulted in significant loss of model fit, then we would have evidence that such a relationship does exist and cannot be ignored. However, adding the constraint did not significantly increase model misfit, $p = .75$, suggesting that there was no meaningful relation between knowledge and the IAT.

Question 3: Are explicit attitudes differentially predictive of original and modified IATs (comparison of Models 2 and 3)? In Study 1, we observed that explicit attitudes were more strongly related to the modified IAT compared to the original IAT. In Model 3, we tested this question by constraining the relationship between the paths between IATs and explicit attitudes to be equal. This resulted in no change in model fit $\Delta\chi^2(1) = 0.1$, $p = .75$, and demonstrates that procedural modifications did not strengthen the attitude-IAT relationship for racial attitudes.

Question 4: Are explicit attitudes related to the IAT at all (comparison of Models 3 and 4)? Our final hypothesis was that there would be a significant and positive relationship between attitude and the IAT. To test this question, we constrained in Model 4 the IAT-attitude relationship to be zero. Compared

to the fit of Model 3 that did not impose such a constraint, Model 4's misfit is substantially greater ($\Delta\chi^2(1) = 31.6$, $p < .001$), *RMSEA* has increased from .06 to .09., and the 95% CI ε_a of Δ does not include .05 (.31-.64). Thus, Model 3, which allows relations between explicit attitude and IATs, is superior to Model 4, and we can conclude that there is a meaningful relationship between explicit attitudes and the IAT.

In summary, four questions were examined with progressive tests of comparative model fit that represented our hypotheses. The data was best fit by Model 3 and a path diagram is presented in Figure 1 as a summary account.[3] For evaluations of Black Americans relative to White Americans, the IAT was unrelated to evaluative knowledge, and was positively related to self-reported attitudes. Those relations did not differ between original and modified IATs. We found no evidence to support the contention that evaluative knowledge is predictive of IAT scores, or that the proposed procedural modifications reduce such a confounding influence.

Study 3 – Food Attitudes

In Study 3, we sought to replicate the observations from Studies 1 and 2 in an investigation of food attitudes inspired by Olson and Fazio's (2004) hypothetical example of people with peanut allergies. We examined attitudes toward peanuts relative to shellfish (another common food allergy), and tested the relationship between different versions of the IAT and self-reported attitudes, behaviors, and knowledge of others' evaluations.

This study was conducted via the Internet and recruited participants through random assignment in a large study pool available at the research portion of the Project Implicit website (https://implicit.harvard.edu/), whether they had a food allergy or not (see Greenwald et al., 2003; Nosek et al., 2002a; and Nosek, 2005 for more information about this virtual research laboratory). We also attempted to supplement this sample with people who had peanut or shellfish allergies. This effort yielded 10 participants with peanut allergies, and they are not examined as a distinct group here (for summary see supplements available at http://briannosek.com/).

Method

Participants

A total of 235 people (average age = 27, *SD* = 11; 69% female) completed the study materials.[4] Of this group, 10 reported having a peanut allergy, 4 reported having a shellfish allergy, 1 reported having allergies to both (coded as having neither for analysis); 187 were White, 10 were Asian, 8 were Black, 9 were Hispanic, 14 were a different ethnicity, and 7 did not report ethnicity. Participants received no compensation for their participation.

Materials

IAT. Design of the IAT followed the procedures described in Study 1. Two evaluative stimulus items (health and sickness) were dropped because of their unique implications for people with allergies toward these foods. Four pictures of shellfish and peanuts served as exemplars for those categories. The data were analyzed with the improved scoring algorithm (Greenwald et al., 2003; five IATs were removed from analysis for too many fast responses, and eight others because of missing data). The IAT effect was coded such that positive values indicated liking for peanuts relative to shellfish.

Explicit measures. Participants completed a questionnaire assessing attitudes, allergies, eating behavior and perceptions of others' evaluations for peanuts and shellfish. Participants rated their food attitudes on semantic differential scales (e.g., disgusting-tasty) and rated their liking and eating behavior. These measures served as indicators of an explicit peanut-shellfish attitude factor. Likewise, on four measures participants rated perceptions of extent to which American culture or the "average person" favored or liked shellfish and peanuts. All measures and descriptive statistics are presented in the Appendix. Positive values indicated greater positivity for peanuts relative to shellfish.

Procedure

After reviewing and agreeing to informed consent procedures, participants performed both versions of the IAT and a brief self-report questionnaire. The order of the questionnaire and implicit measures was randomized, as was the order of the two IATs.

Results and Discussion

Unique Attitudinal Variation in the Original and Modified IATs

We first sought to demonstrate that the original and modified IATs each contained unique attitudinal variation. A structural model in which explicit attitude was regressed on original and modified IATs showed that both the modified ($\beta = .32$, $p < .001$) and original ($\beta = .36$, $p < .001$) IATs contained significant unique attitudinal variation, replicating Study 1 with food attitudes.

Relations between the IAT and Self-Reported Attitudes and Evaluative Knowledge

We examined the strength of the relationship between evaluative knowledge and the two versions of the IAT. If the original IAT was influenced by knowledge about others' evaluations of these foods, then these factors should correlate positively. Also, if the modified IAT reduces or removes the influence of extra-personal knowledge, then it should show significantly weaker correspondence with the knowledge items.

We followed the hypothesis testing approach outlined in Study 2 to evaluate the relations between the IAT, explicit attitudes, and evaluative knowledge. As before, a sequence of hypotheses were evaluated by comparing model fit statistics that directly examined specific hypotheses in the context of a full model of relations. The sequence of models is identical to Study 2 and is summarized briefly (full model fit information is available in Table 1).

The initial model allowing all parameters to vary freely fit the data reasonably well ($\varepsilon_a = .061$). Evaluative knowledge was not differentially related to original and modified versions of the IAT, $\Delta\chi^2(1) = 0$, $p = .99$ (comparing Models 0 and 1), suggesting that the procedural modifications had no effect on the relations between knowledge and the IAT. Further, fixing the relationship between evaluative knowledge and the IAT to zero had minimal impact on model fit, indicating that evaluative knowledge was unrelated to IAT scores, $\Delta\chi^2(1) = 0.8$, $p = .37$ (comparing Models 1 and 2). The IAT-attitude relationship did not vary between IAT versions, $\Delta\chi^2(1) = 0$, $p = .99$ (comparing Models 2 and 3). Finally, the relationship between explicit attitudes and the IAT was positive and significant; fixing that relationship to zero significantly increased model misfit, $\Delta\chi^2(1) = 115.6$, $p < .0001$ (comparing Models 3 and 4).

In summary, these progressive model tests replicated prior observations in the domain of food attitudes. Figure 2 shows Model 3, representing the best fitting model. Again, the IAT was not related to evaluative knowledge and was positively related to self-reported attitudes. The procedural modifications to the IAT did not moderate either of these relations. We found no evidence for an extra-personal knowledge confound in the IAT, and no evidence to support the contention that the proposed IAT procedural modifications help to reduce a confounding influence.

Studies 4, 5, and 6

In the first three studies we demonstrated that (a) the original and modified IATs each capture unique attitudinal variation (Studies 1 and 3), (b) the IAT did not correspond with evaluative knowledge (Study 2 and 3), (c) the IAT corresponded with explicit attitudes (Studies 1-3), and (d) the modified IAT was more strongly related to explicit attitudes than was the original IAT in one of three topical domains (political attitudes; Study 1). These results suggest that the proposed IAT procedural modifications are altering measurement, but that the prevailing interpretation that it does so by reducing an extra-personal knowledge confound is unsupported. Yet, the fact that IAT-attitude correspondence increased with the modifications begs the question, if the procedures are not removing confounding variance from the IAT (due to knowledge or any other source), why was this difference observed? In Studies 4-6, we sought to replicate and extend the conclusions of these studies, and also test our hypothesis that the IAT procedural variations designed to remove a confound actually creates one instead.

Identification of a Confounding Factor in the Modified IAT

Olson and Fazio (2004) introduced two procedural modifications to the IAT intended to remove a presumed influence of extra-personal knowledge: (1) error feedback was removed, and (2) the category labels were changed to “I like” and “I dislike” for judgments of evaluative exemplars. In the two critical portions of the IAT participants categorize stimulus items representing one of four categories (e.g., John Kerry, George Bush, I like, and I dislike) using two response keys. In one case, the response keys are labeled such that items representing “John Kerry” and “I like” are categorized with one key, and items representing “George Bush” and “I dislike” are categorized with the other key. In the other case, the

response keys are labeled such that items representing “John Kerry” and “I dislike” are categorized with one key, and items representing “George Bush” and “I like” are categorized with the other key.

A critical procedural demand is that participants perform one of two distinct judgment tasks depending on the stimulus dimension. For “I like-I dislike” exemplars, categorization is based on evaluative features (liked or disliked). For “John Kerry-George Bush” exemplars, categorization is based on category membership (Kerry or Bush), not evaluative features (whether the candidates are liked or disliked).

We hypothesized that removing error feedback and using labels “I like” and “I dislike” may increase confusion between judgment tasks, such that target concept exemplars are more likely to be categorized based on evaluative features contrary to the IAT procedural requirements. That is, the procedural innovations may subtly encourage a confounding influence in measurement – the explicit evaluation of target concepts.

In the original IAT design, error feedback reminds respondents of the proper judgment task. For example, imagine that a Republican is performing a political attitude IAT with the key assignment requiring sorting “George Bush” with “Unpleasant” to the left and “John Kerry” with “Pleasant” to the right. If she sees the item “John Kerry” and hits the left key for “Unpleasant”, the error feedback reminds her that “John Kerry” is supposed to be categorized as a “John Kerry” or “George Bush” and not evaluated as “Pleasant” or “Unpleasant”. In other words, the error feedback reinforces the sorting rules established by the category labels. If the error feedback is removed, she may not notice that she recoded the task instructions and might continue to categorize “John Kerry” as “Unpleasant” and ‘George Bush’ as “Pleasant” violating the procedural requirements. As a consequence, the error rate for “John Kerry” and “George Bush” judgments would increase when the response key assignments are mismatched with her explicit evaluations of John Kerry and George Bush.

Using the labels “I like” and “I dislike” might further exacerbate task recoding by encouraging subjects to think about their explicit evaluations of stimulus items (both target concepts and evaluative attributes). Indeed, this is the purpose of these labels, as Olson and Fazio (2004) state: “changing the

labels from '[Un]pleasant' to 'I [don't] like' may be enough to direct participants to construe the items presented in terms of their own attitudes and to reduce the influence of extrapersonal associations used to solve the IAT's mapping problem" (p. 658). Our concern, however, is that this change may unintentionally extend beyond the evaluative stimulus items and result in explicit evaluation of the categorical items as well.

Our task-recoding hypothesis leads to a specific prediction about the pattern of errors that should emerge because of the procedural modifications. In the modified IAT, task recoding should be increasingly likely for the target items (e.g., John Kerry/George Bush) but the evaluative items (e.g., I like/I dislike) may be relatively unaffected. In other words, subjects should show an increased likelihood of evaluating John Kerry and George Bush items as "I like" or "I dislike".

Whether this task recoding increases or decreases categorization errors depends on the response key assignments in the IAT's two critical blocks. If anything, task recoding should decrease errors when the explicit evaluation of target categories matches the key assignment (e.g., for Republicans, "John Kerry" with "I dislike" on one key and "George Bush" with "I like" with the other) because it simplifies the task from a 4-category judgment task (John Kerry, George Bush, I like, I dislike) to a 2-category task (I like, I dislike). On the other hand, task recoding should increase error rates when the explicit evaluation of target categories mismatches the key assignment (e.g., for Republicans, "John Kerry" with "I like" with one key and "George Bush" with "I dislike" with the other). So, evidence for task recoding will be observed if there is a magnitude increase in the absolute difference in error rates between the two critical conditions for the IAT for the target concepts.

We tested this task-recoding hypothesis in Studies 4-7. In addition to the two versions of the IAT examined thus far, we introduced a 'hybrid' version of the IAT that retained the 'Pleasant/Unpleasant' labels but removed error feedback. This hybrid version should help to determine whether error feedback and label changes each contribute to increased influence of task recoding in the modified IAT.

Replicating and Extending Conclusions from Previous Studies

For a previous version of this paper, reviewers raised a number of concerns about potential limitations of inferences drawn from Studies 1-3 such as: (a) *power* – even though self-reported attitudes related more strongly to the IAT than did knowledge, perhaps the lack of relationship between knowledge and the IAT was a consequence of insufficient power (however, note that *N*s ranged from 82 to 235, and power to detect an *r* of .30 with a 2-tailed test at $\alpha = .05$ ranged from .78 to .997), (b) *representation of knowledge* – perhaps the assessment of evaluative knowledge was too narrow, and a more diverse assessment of knowledge would show relations to the IAT, (c) *validity of evaluative knowledge* – knowledge items used in Studies 1-3 related to each other but no measures were included to show their criterion validity, so perhaps evaluative knowledge was just poorly assessed, and (d) *multiple measures* – the primary measurement comparisons were within-subjects, and perhaps completing multiple measures disrupted their proper assessment and the effects would be different in a between-subjects design.

Studies 4-6 were designed to address all of these concerns. In particular, (a) Studies 4-6 have large samples resulting in high powered tests of the hypotheses (Study 4 $N = 1{,}124$; Study 5 $N = 739$; Study 6 $N = 1{,}100$), (b) Studies 4-6 introduced a wider range of evaluative knowledge items to ensure that the null effects in the earlier studies were not due to possible narrowness in knowledge assessment, (c) Studies 4-6 include criterion variables that should be predicted by knowledge to ensure that the measures are valid, and (d) Studies 4-6 were conducted with between-subjects manipulations of implicit measures to eliminate concerns about order effects. Following Studies 4-6, Study 7 ($N = 12{,}152$) established the generality of these observations across 58 different content domains. Because Studies 4-6 used similar methods, they are described together with results and discussion following.

Method for Studies 4, 5, and 6

Participants

Participants were recruited through the research Web site of Project Implicit. After participants register an identity at Project Implicit, they are randomly assigned to a study from the study pool each time they log-in to Project Implicit. Participants are not assigned to the same study more than once.

For Study 4, 1,124 participants provided data for analysis after dropping 4 for not following task instructions, 19 for too many fast responses, and 246 interrupted sessions with no IAT data. For Study 5, 735 participants provided data for analysis after dropping 1 for not following task instructions, 3 for too many fast responses (>10%), and 104 interrupted sessions with no IAT data. For Study 6, 1,197 participants provided data for analysis after dropping 1 for not following task instructions, 28 for too many fast responses, and 237 interrupted sessions with no IAT data (for more on web procedures see Nosek et al., 2002a; Nosek, 2005).

Materials

While the content varied, the form of the materials was constant across Studies 4, 5, and 6. The content paralleled the domains examined by Olson and Fazio (2004): Study 4 concerned attitudes toward John Kerry compared to George Bush, Study 5 concerned attitudes toward Black compared to White people, and Study 6 concerned attitudes toward Candy Bars compared to Apples. Stimulus materials appear in the Appendix.

IAT. Three versions of the IAT were created for each content domain. The original version followed the description in Study 1 with error feedback and the evaluative labels "Pleasant/Unpleasant" as distinguishing features. The modified version also followed the description in Study 1 with no error feedback and evaluative labels "I like/I dislike" as distinguishing features. Finally, a hybrid version was introduced that did not provide error feedback and used the evaluative labels "Pleasant/Unpleasant". The IAT showed good internal consistency (Study 4 – politics, $\alpha = .90$; Study 5 – race, $\alpha = .88$; Study 6 – food, $\alpha = .86$).

Explicit attitudes, evaluative knowledge, and knowledge criterion variables. For each study a collection of attitude, knowledge, and knowledge criterion items were administered (see the Appendix for a description of items). Items were similar to previous studies, though additional knowledge questions were administered to broaden representation of that factor. Also, criterion variables for knowledge items in each content domain were identified to demonstrate predictive validity of evaluative knowledge.

Procedure

The procedure was the same for Studies 4, 5, and 6. After being randomly assigned to the study and giving informed consent, subjects completed an IAT and a short questionnaire. Subjects were randomly assigned to complete one of three versions of the IAT (original, hybrid, modified). Presentation of the IAT and questionnaire was randomized across subjects. Also, item order in the questionnaire was randomized.

Results and Discussion for Study 4 (Political Attitudes)

Criterion Validity of Evaluative Knowledge

One concern with the previous studies was that the lack of relationship between the IAT and knowledge could have been due to weaknesses in the measurement of evaluative knowledge. In Study 4, two criterion variables were included that we hypothesized to be related to evaluative knowledge. All six knowledge items were positively related to participants' perceptions of who would be elected if the election were held 'today' – the day of their participation (*r*s = .27 - .54, median *r* = .44, *p*s < .0001), and who would be elected in the upcoming November election (*r*s = .23 - .54, median *r* = .44, *p*s < .0001). Simultaneously, the knowledge items were less related to participants' report of who they would vote for themselves (*r*s = .05 - .36, median *r* = .30, *p*s = .12 to <.0001).

The knowledge items were positively inter-correlated (*r*s = .33 - .75, median *r* = .52), and exploratory principle component factor analysis yielded only one eigenvalue ≥ 1.0 (3.5), indicating that a single latent factor is sufficient to account for relations between these items. Additional analyses conducted on each knowledge item separately, instead of using the latent factor in the models were consistent with the effects reported below.

Relations between the IAT, Explicit Attitudes and Evaluative Knowledge

We examined our hypotheses concerning the relations between attitudes, knowledge and the IAT following the comparative model fitting approach used in Studies 2 and 3. However, since the IAT version was manipulated between-subjects in Studies 4-6, rather than within subjects, a multiple-group (two) structural modeling approach was used. Model fit results are presented in the first panel of Table 2.

For simplicity and consistency with Studies 2 and 3, the hybrid IAT was not included in these models. It is featured in the following section examining the task recoding confound.

Question 0.1: Can factorial invariance be assumed for measurement models across groups? A confident comparison of latent factors between groups (original and modified IATs) requires that the underlying measurement models show factorial invariance (Horn & McArdle, 1992; Meredith, 1964; Thurstone, 1947), that is, that the pattern of factor loadings is similarly across groups. Factorial invariance was examined by constraining the factor loadings equal across groups and comparing the fit with a model in which all loadings varied freely. Adding the loading constraint resulted in a non-significant change, $\Delta\chi^2(9) = 16.5$, $p = .06$ and a slight improvement in overall model fit (95% CI ε_a of Δ= .000, .063), demonstrating that the assumption of invariance across groups is justified.

Question 0.2: Is the relationship between explicit attitudes and evaluative knowledge different across conditions (comparison of Models 0.1 and 0.2)? Whether participants completed the original or modified IAT should have no impact on the relationship between evaluative knowledge and explicit attitudes. This preliminary prediction was tested by constraining the attitude-knowledge relationship equal between groups. Consistent with our hypothesis, this resulted in minimal loss of model fit compared to the unconstrained model, $p = .53$.

Question 1: Is knowledge differentially predictive of original and modified IATs (comparison of Models 0.2 and 1)? We hypothesized that procedural modifications to the IAT would not affect the relationship between the IAT and evaluative knowledge. If that is true, then constraining the knowledge-IAT relations to be equal across IAT conditions should have little impact in overall model fit. Adding such a constraint (Model 1) resulted in a trivial change in χ^2 relative to Model 0.2, $\Delta\chi^2(1) = 0.5$, $p = .48$. This suggests that the procedural modifications had no effect on the IAT's relation with evaluative knowledge.

Question 2: Is evaluative knowledge related to the IAT at all (comparison of Models 1 and 2)? We predicted that evaluative knowledge would have no meaningful relationship to IAT scores. This question was tested by fixing for both groups the relationship between evaluative knowledge and the IAT

to zero. If constraining this relationship to be zero (Model 2) results in a meaningful decrement in model fit, then we would have evidence that such a relationship does exist and should not be ignored. The difference in χ^2 between Models 2 and 1 is statistically significant ($\Delta\chi^2(1) = 3.9$, $p = .05$), but trivial in the context of the models' complexity and sample size. This assertion is supported by observing that the *RMSEA* for Model 2 was unchanged ($\varepsilon_a = .060$), and the 95% CI for the *RMSEA of the change* includes .05 (.00-.15). This indicates that the fits of these models are not substantively different and that the gain in parsimony from the added constraint compensated for the minor loss of fit. Further, the relationship between knowledge and the IAT was actually in the opposite direction predicted by prior theories (Karpinski & Hilton, 2001; Olson & Fazio, 2004). That is, people who believed that others had a greater preference for Bush relative to Kerry actually showed slightly greater preference for Kerry relative to Bush on the IAT ($\beta = -.07$; see Figure 3). Despite the minimal change in fit, the small magnitude of effect, and the opposite direction from prediction, this is the first inkling of there being any relationship between knowledge and the IAT. So, to give the hypothesis that knowledge and the IAT are related every chance to persist, we did not retain the zero-constraint on the IAT-knowledge relationship.

Question 3: Are explicit attitudes differentially predictive of original and modified IATs (comparison of Models 1 and 3)? The previous studies showed inconsistency in whether the procedural modifications to the IAT significantly altered its relationship with self-reported attitudes. Study 1 showed a difference in correlation strength, but Studies 2 and 3 did not. In Model 3, we constrained the relationship between the IAT and explicit attitude to be equal across IAT conditions and again observed, despite the small, but statistically significant change in χ^2 compared to the less constrained Model 2 ($\Delta\chi^2(1) = 7.5$, $p = .01$) that these models are not substantially different ($\varepsilon_a = .061$ for Model 3, .060 for Model 1; and 95% CI ε_a of Δ includes .05. However, because the difference in correlation strength between explicit attitudes and the IAT versions is critical to the hypothesis proposed by Olson and Fazio (2004), we retained the less constrained Model 2. The procedural modifications did affect the relationship between the IAT and explicit attitude for these political attitudes, and the IAT-attitude relationship was stronger for the modified IAT than for the original IAT (see Figure 3).

Question 4: Are explicit attitudes related to the IAT at all (comparison of Models 1 and 4o and 4m)? Because the original and modified IATs exhibited different relations with self-reported attitudes, testing the magnitude of the relationship between attitudes and the IAT were conducted separately for each IAT version. Model 4o fixed the relationship between explicit attitudes and the original IAT to zero and was compared to Model 1 to test whether defining a null attitude-original IAT relationship resulted in an important loss of fit. Model 4m allowed for the same comparison but with the explicit attitude-modified IAT relationship fixed to zero.

The results presented in Table 2 make clear that in both cases, fixing the IAT-attitude relationship to zero had substantial deleterious effects on the quality of model fit (original: $\Delta\chi^2(1) = 165$, $p < .0001$; modified: $\Delta\chi^2(1) = 260$, $p < .0001$). The ε_a was likewise negatively affected by adding these constraints and the 95% CIs for *RMSEAs of the changes* no longer include .05. We can thus conclude that there is a meaningful relationship between explicit attitudes and the IAT, both in its original and modified form.

Model 1 is presented in Figure 3 as the summary account of the results. For evaluations of George Bush relative to John Kerry, the IAT was weakly but negatively related to evaluative knowledge, and was strongly and positively related to self-reported attitudes. Also, replicating earlier work, the procedural modifications resulted in somewhat stronger positive correspondence with self-reported attitudes (β for original IAT = .71, modified IAT = .79). Consistent with the prior studies, we found no evidence to support the contention that the proposed procedural modifications reduce a confounding influence of evaluative knowledge, even in this case where a weak knowledge-IAT relationship was observed.

Task Recoding in the Modified IAT

A second goal of Study 4 was to examine our hypothesis that the procedural changes for the modified IAT introduced a task recoding confound in which the target concepts (Bush-Kerry) are increasingly likely to be explicitly evaluated rather than categorized based on the intended nominal feature (person identity). We hypothesized that the absolute difference in error rates between critical blocks would be higher in the hybrid IAT (no error feedback, but retaining the "Pleasant/Unpleasant"

evaluative labels) compared to the original IAT for the political target items. Further, we hypothesized that the absolute difference in error rates would be higher still in the modified IAT compared to the hybrid IAT because the introduction of the labels "I like" and "I dislike" would further encourage explicit evaluation of the political targets.

The critical test of the hypothesis regarding differences in error rates is presented in the first panel of Figure 4. The pattern corresponds to prediction with increased differential error rates for the political targets for the hybrid over the original, and the modified over the hybrid, IAT. Simultaneously, there was little variation in differential error rates for the evaluative targets. Reflecting the observed pattern, the interaction effect between targets and IAT version was significant $F(2, 1059) = 9.97$, $p < .0001$, and follow-up tests showed that for political targets, the modified IAT elicited slightly but not significantly greater differential error rates than the hybrid IAT, and both elicited greater differential error rates than the original IAT. Only minor differences were observed between conditions for the evaluative targets.

In sum, the data support our hypothesis that removing error feedback and introducing "I like" and "I dislike" labels encourages evaluative task recoding of target concepts. In this example, that effect is driven predominately by the removal of error feedback. The IAT procedural modifications do not remove a confound, they introduce one.

Summary

Study 4 replicated and extended the findings from Studies 1-3. The IAT was demonstrated to be related to explicit attitudes and weakly, but negatively, to evaluative knowledge. Replicating previous studies, the IAT modifications were shown to strengthen the attitude-IAT relation but did not affect the knowledge-IAT relation. And, new in Study 4, we demonstrated that the IAT modifications introduced a task recoding confound in which the target concepts were explicitly evaluated.

Results and Discussion for Study 5 (Racial Attitudes)

Studies 5 and 6 are direct replications of Study 4 in distinct content domains. Analysis followed the approach described in Study 4, and results are presented briefly to conserve space.

Predictive Validity of Evaluative Knowledge

Confirming that there is meaningful variability in evaluative knowledge, all six knowledge items were significantly positively correlated with estimates of employers' preferences for hiring White over Black job candidates, and likelihood estimates of group members being targets of discrimination. Those who perceived others' to have stronger pro-White preferences predicted more pro-White hiring practices (*r*s range = .25 - .44; median *r* = .34; *p*s < .0001), but self-reported attitudes were unrelated to perceived hiring practices (*r*s = -.05, -.06; *p*s = .13, .16). Also, those who perceived others' to have stronger pro-White preferences predicted greater likelihood of Blacks being discriminated against compared to Whites (*r*s range .13 - .45; median *r* = .33; *p*s < .0005), but self-reported attitudes were negatively related to predicted discrimination rates (*r*s = -.17, -.15; *p*s < .0001). In sum, evaluative knowledge shows criterion validity for judgments of others' behavior.

An exploratory factor analysis of the knowledge items revealed two eigenvalues ≥ 1.0 (2.8, 1.2), leading to a correlated, two-factor model (Factor 1: liking and preferences of average and most people; Factor 2: warmth and historical and social portrayals; see Appendix) of evaluative knowledge As before, analyses with the individual manifest knowledge items were consistent with the reported effects unless otherwise noted.[5]

Relations between the IAT, explicit attitudes, and evaluative knowledge

To examine the relations between the IAT, explicit attitudes, and evaluative knowledge, Study 5 followed the same series of hypothesis tests described in Study 4. The results are summarized in the second panel of Table 2.

The initial model allowing all parameters to vary freely was a good fit to the data (ε_a = .036) and comparison to Model 0.1 supports an assumption of factorial invariance, $\Delta\chi^2(7) = 6.8$, $p = .45$. Further constraining the relationship between explicit attitudes and evaluative knowledge to be equal across IAT conditions had no effect on overall model fit, $p = .64$ (comparing Models 0.1 and 0.2).

Concerning the substantive hypotheses, evaluative knowledge was not differentially related to original and modified versions of the IAT, $\Delta\chi^2(2) = 1.1$, $p = .58$ (comparing Models 0.2 and 1), suggesting that the procedural modifications had no effect on the relations between knowledge and the

IAT. Fixing the relationship between evaluative knowledge and the IAT to zero had a small but significant impact on model fit indicating that evaluative knowledge was related to IAT scores, $\Delta\chi^2(2) = 6.0$, $p = .05$ (comparing Models 1 and 2). As in Study 4, despite the significant change in χ^2, the change in ε_a was minimal and the 95% confidence interval included .05 (see Table 2), again suggesting that statistical significance was a consequence of large sample size. Examination of beta weights revealed that one relation was positive and the other negative. Further, follow-up tests on the individual knowledge items showed that only one (historical portrayals) was significantly related to the IAT ($p = .051$) and in the wrong direction ($\beta = -.14$). Even so, the relation was retained to promote the best opportunity for a knowledge-IAT relation to persist.

Replicating the observation for racial attitudes in Study 2, the IAT-attitude relationship did not vary between IAT versions, $\Delta\chi^2(1) = .4$, $p = .53$ (comparing Models 1 and 3). Finally, the relationship between explicit attitudes and the IAT was positive and significant and fixing that relationship to zero significantly increased model misfit ($\Delta\chi^2(1) = 30.6$, $p < .0001$), the change in *RMSEA* was substantial and its 95% confidence interval did not include .05 (comparing Models 3 and 4).

The summary model is represented in Figure 5. For evaluations of Blacks relative to Whites, the IAT was positively related to self-reported attitudes and showed slight positive and negative relations with evaluative knowledge. The procedural modifications to the IAT did not moderate either of these relations. While showing a significant IAT-knowledge relation, these data were not very reassuring for the knowledge confound hypothesis. The knowledge relation was in the opposite direction from prediction for one factor, and the effect was so small that model fit indices suggested that it was meaningless. The lack of adjustment to the significance level in the context of the large sample sizes may be the culprit. We return to this issue in Study 7 and the General Discussion.

Task Recoding in the Modified IAT

We also replicated our observation that the procedural changes for the modified IAT introduced a task recoding confound in which the target concepts (Black-White) would be more likely to be explicitly evaluated rather than categorized based on the intended nominal feature (race). The critical results are

presented in the second panel of Figure 4. Replicating Study 4, the pattern corresponds to prediction with increased differential error rates for the racial targets for the hybrid over the original, and the modified over the hybrid, IAT. Simultaneously, there was little variation in differential error rates for the evaluative targets. The interaction effect between targets and IAT version was significant $F(2, 715) = 33.02$, $p < .0001$), and follow-up tests showed that for racial targets, the modified IAT elicited significantly greater differential error rates than the hybrid IAT, which itself elicited greater differential error rates than the original IAT. Only minor differences were observed between conditions for the evaluative targets. In this case, both removing errors and introducing the new evaluative labels contributed to the increase in task recoding. In sum, the data conform neatly to the hypothesis that removing error feedback and introducing "I like" and "I dislike" labels encourages evaluative task recoding of target concepts.

Results and Discussion for Study 6 (Food Attitudes)

Predictive Validity of Evaluative Knowledge

Confirming that there is meaningful variability in evaluative knowledge, all six evaluative knowledge items were significantly correlated with perceptions of consumer purchasing behavior of apples compared to candy bars, and perceptions of which item the most people would choose if given a choice to eat. Those who perceived the culture to have stronger candy bar preferences predicted more purchasing of candy bars compared to apples in stores (rs range = .07 - .22; median r = .18; ps = .02 - < .0001), but self-reported attitudes were barely related to purchasing estimates (rs = .06, .06; ps = .05, .05). Also, those who perceived others' to have stronger candy bar preferences predicted more frequent candy bar selection compared to apple selection by others when given a choice (rs range .12 - .45; median r = .25; ps < .0001), but self-reported attitudes were unrelated to such estimates (rs = .03, .04; ps = .18, .33). In sum, evaluative knowledge shows criterion validity for perceptions and predictions of others' food-related behavior.

The knowledge items were positively correlated with one another (rs = .05 - .68, median r = .17), and preliminary analysis indicated some heterogeneity such that the best latent model of evaluative

knowledge involved two correlated factors (eigenvalues ≥ 1.0: 2.3, 1.6; Factor 1: liking and warmth of society and estimates of average and most people's preferences; Factor 2: historical and social portrayal). A two-factor representation of evaluative knowledge was used for the subsequent analyses. The same analyses conducted on each of the individual manifest knowledge items were consistent with the reported results.

Relations between the IAT, Explicit Attitudes, and Evaluative Knowledge

In Study 6 the target concepts of interest were evaluations of food (Candy Bars versus Apples). The results are summarized in the third panel of Table 2. Again, the initial model allowing all parameters to vary freely fit the data reasonably well ($\varepsilon_a = .056$) and constraining the factor loading equal across groups resulted in minimal loss of fit, $\Delta\chi^2(7) = .9$, $p = .99$ (comparing Models 0 and 0.1). Also, constraining the relationship between explicit attitudes and evaluative knowledge to be equal across IAT conditions had no effect on overall model fit, $\Delta\chi^2(3) = 1.9$, $p = .59$ (comparing Models 0.1 and 0.2).

More critically, evaluative knowledge was not differentially related to original and modified versions of the IAT, $\Delta\chi^2(2) = 1.3$, $p = .52$ (comparing Models 0.2 and 1) suggesting that the procedural modifications had no effect on the relations between knowledge and the IAT. Fixing the relationship between evaluative knowledge and the IAT to zero had no impact on model fit, $\Delta\chi^2(2) = 1.7$, $p = .43$ (comparing Models 1 and 2).

Previous studies demonstrated inconsistency in whether the IAT modifications would elicit stronger correspondence with explicit attitudes. Constraining the attitude-IAT relationship to be equal across IAT versions for Candy Bar-Apple attitudes resulted in a significant change in χ^2, $\Delta\chi^2(1) = 4.5$, $p = .03$ (comparing Models 2 and 3). However, like Study 4, there was no change in overall model fit ($\varepsilon_a = .051$). Even so, because this difference is central to the hypothesis that the IAT modifications reduce the influence of confounding variance, we did not force this equality constraint.

Because the attitude-IAT relationship was not constrained equal across groups, two separate models tested setting the attitude-IAT relationship to zero for the original and modified IATs. In both cases, constraining the attitude-IAT relationship to zero resulted in a substantial increase in model misfit

(Original IAT only: $\Delta\chi^2(1) = 43.3$, $p < .0001$; Modified IAT only: $\Delta\chi^2(1) = 80.5$, $p < .0001$; comparing Model 2 with 4o and 4m), and the 95% confidence interval of the ε_a of change clearly shows a decline in fit that does not justify adding this constraint. This indicates that both versions of the IAT were related to self-reported attitudes.

These effects are summarized graphically by Model 2 in Figure 6. For evaluations of Candy Bars relative to Apples, the IAT was not related to evaluative knowledge and was related to self-reported attitudes. The procedural modifications to the IAT altered the attitude-IAT relationship such that the modified IAT showed somewhat stronger relations to self-reported attitudes. Despite high power and a heterogeneous array of evaluative knowledge assessments, we found no evidence for an evaluative knowledge confound in the IAT, and no evidence to support the contention that the proposed procedural modifications to the IAT help to reduce such a confounding influence.

Task Recoding in the Modified IAT

We again examined our hypothesis that the procedural changes for the modified IAT introduced a task recoding confound in which the target concepts (Candy Bar-Apple) would be explicitly evaluated rather than categorized based on the intended nominal feature (food type). The results are presented in the third panel of Figure 4. The pattern corresponds to prediction with increased differential error rates for the food targets for the hybrid over the original, and the modified over the hybrid, IAT. Simultaneously, there was little variation in differential error rates for the evaluative targets. Reflecting the observable pattern, the interaction effect between targets and IAT version was significant $F(2, 1116) = 10.5$, $p < .0001$), and follow-up tests showed that for food targets, the modified IAT elicited slightly but not significantly greater differential error rates than the hybrid IAT, while both elicited greater differential error rates than the original IAT. No differences were observed between conditions for the evaluative targets.

Study 7

Studies 4-6 added to the demonstrations in Studies 1-3 by showing that the IAT was robustly related to explicit attitudes across content domains, while it showed weak to absent relations with

evaluative knowledge. Further, a task recoding confound was identified that resulted from the procedural modifications to the IAT.

The previous studies suggest that the differences in IAT-attitude correlations between IAT versions may be rather weak and inconsistently observed across content domains. A significant difference was observed for 3 studies (1, 4, 6; the same content domains reported by Olson and Fazio, 2004) and no difference was observed in the other three (2, 3, 5). Also, knowledge was slightly related to the IAT in two studies (4, 5), though not in three others (2, 3, 6). The fact that the IAT-knowledge relation was unstable, often in the wrong direction, and so weak that it did not affect model fit, raises suspicions that the significant effects were due to sample size and Type I error. It is possible, however, that a relationship between knowledge and the IAT will appear in some content domains and not others.

In Study 7, we sought to investigate the generality of our observations across a wide variety of topics. In this case, we conducted a large study ($N > 12{,}000$) in which we examined task recoding and the relations among the IAT, explicit attitudes, and evaluative knowledge across 58 different content domains.

Method

Study 7 was nearly identical to the study described by Nosek (2005) that examined moderators of the relationship between implicit and explicit preferences. The single change was to randomly assign participants to complete the original or modified IAT. All other procedures and materials were identical and described completely in Nosek (2005). Only those measures that were critical to the present investigation are described here.

Participants

A total of 12,972 tasks were completed by 7,401 volunteer participants. After login, participants were randomly assigned to one of the 58 topical domains.[6] Of the 7,401 participants, 5,023 (68%) completed just one task. Using only the first study completion for each participant for analysis does not substantively influence the effects reported here (see also Greenwald et al., 2003; Nosek et al., 2002a).

Of the 7,401 participants who reported demographic information (98% response rate) the following was observed: 68% female, 32% male; 1% American Indian, 5% Asian, 7% Black, 5% Hispanic, 73% White, 1% Biracial (Black-White), 4% Multiracial, and 4% Other; 20% Conservative, 31% Neutral or Moderate, and 49% Liberal; and, the average participant was born in 1974 (i.e., ~30 years old; *SD* = 11.6 years). Following data cleaning (dropping tasks with missing data or when >10% of response latencies were shorter than 300ms; see Greenwald et al., 2003), a total of 12,152 usable completed tasks remained.

Materials

IAT. Design of modified and original IATs followed the procedures described in Study 1. The object pairs and stimulus exemplars were the same as those described in Nosek (2005). IAT scores were calculated such that positive values indicated an implicit preference for the concept implicitly preferred on average. The explicit attitude and evaluative knowledge measures were similarly scaled.

Self-reported attitudes. Relative explicit attitudes were assessed by calculating the difference between feelings of warmth ratings for the two attitude objects. Feelings of warmth were assessed on a 9-point scale (1-9) in which 1 indicated very cold feelings and 9 indicated very warm feelings. As such, the difference score between the two ratings had a range of possible values from -8 to +8 with 0 indicating explicit attitude indifference (no relative explicit preference for one attitude object over the other).

Evaluative knowledge. Evaluative knowledge was represented with estimates of how the average person felt about the concepts using the same warm-cold rating scale at the attitude measure. Like the relative explicit attitude rating, the difference between the average person ratings indicated the perceived relative preference of others' for one attitude object over the other.

Procedure

The study was administered via the research website for Project Implicit (version 1.6; https://implicit.harvard.edu) between October 13, 2003 and September 17, 2004. Once randomly assigned to a study, participants completed explicit measures and either the original or modified IAT in a

randomized order. Explicit measures were presented on a single webpage, but their order was randomized. IAT procedures followed those described in Study 1.

Results and Discussion

Original and Modified IATs Relation with Explicit Attitudes and Evaluative Knowledge

Simultaneous regressions examined attitudes and knowledge as predictors of IAT scores separately for the original and modified IATs for each of the 58 topical domains. Beta weights for the predictors are presented in Table 3 under the 'Regression Summary' heading. For the original IAT, explicit attitudes significantly predicted IAT scores in 46 of the 58 topical domains (79%), and all significant relations were positive (average $\beta = .33$). Evaluative knowledge significantly predicted IAT scores in 2 of the 58 domains (3.4%; average $\beta = .04$) with one being significantly positive (Meg Ryan-Julia Roberts) and the other negative (Christian-Jewish).

For the modified IAT, explicit attitudes significantly predicted IAT scores in 44 of the 58 topical domains (76%), and all significant relations were positive (average $\beta = .34$). Also, evaluative knowledge significantly predicted modified IAT scores in 4 of the 58 domains (6.9%; average $\beta = -.03$) though two were significantly negative (Skirts-Pants, Tea-Coffee) and two were significantly positive (American-Canadian, Meg Ryan-Julia Roberts). In short, explicit attitudes were strong predictors of IAT scores and evaluative knowledge was, at best, a weak predictor of IAT scores whether using the original or modified versions of the task. [7]

We further examined this relationship by comparing original and modified IATs in single analysis with 58 independent simultaneous regressions of condition (original or modified), explicit attitude, evaluative knowledge, and interaction terms of those three factors predicting IAT scores. If attitudes and knowledge are more strongly related to one version of the IAT compared to the other, we would observe significant condition X attitude or condition X knowledge interaction terms. Of the 58 simultaneous regressions, in five (7.7%) the condition X attitude interaction was significant ($p < .05$), and in three (5.2%) the condition X knowledge interaction was significant.

For four of the significant condition X attitude interactions, the modified IAT was more strongly related to self-reported attitudes than was the original IAT (Short people-Tall people, Rich people-Poor People, McDonalds-Burger King, Future-Past), and for one, the original IAT was more strongly related to self-reported attitudes than was the modified IAT (Thin people-Fat people). For all three of the significant condition X knowledge interactions (California-New York, Nerds-Jocks, Future-Past), the modified IAT was somewhat (but non-significantly) negatively related to evaluative knowledge while the original IAT was somewhat (but non-significantly) positively related to knowledge.

In summary, the IAT was consistently, positively related to self-reported attitudes, and consistently unrelated to evaluative knowledge. The procedural modifications to the IAT had modest effects on its relation to attitudes and knowledge. The modifications did slightly, though erratically, increase the relation between attitudes and the IAT, but had little effect on the relation between knowledge and the IAT. Significant relations with knowledge were observed in both directions at frequencies consistent with expectations of Type I error at $\alpha = .05$. We next examined whether task recoding increased as a consequence of the procedural modifications and whether that change was sufficient to account for the small increase in correspondence between the IAT and explicit attitudes.

Task Recoding

As in Studies 4-6, task recoding was evidenced by a differential increase in categorization errors for target concepts (e.g., Black and White faces) between response blocks. If participants errantly categorize target concepts based on their judged evaluative properties (liked or disliked) rather than their category identities, then the error rate when the category and judged evaluation are on same response key will be lower than when they are on different response keys. We had hypothesized that removing error feedback and using evaluative labels "I like/I dislike" would increase the frequency of this recoding.

The far right columns of Table 3 present the absolute mean difference in categorization errors of target concept items for the original and modified IATs. Also, the last column presents an effect size (Cohen's *d*) estimate of the difference between the differential error rates, and an '*' indicates that the difference was significantly different from zero ($p < .05$). Remarkably, without exception, the differential

error rate was greater in the modified IAT than the original IAT, though significantly so for only 38 of the 58 topical domains (66%; average d = .32). That 66% of the significance tests were significant is consistent with a power analysis for detecting a d = .3 effect for a sample with the average characteristics of these data (between-groups comparison, non-homogeneous variances, average N ~ 200, power = .60), suggesting that task recoding is a recurrent and stable result of the procedural modifications to the IAT.

In the previous section, we reported that five of the 58 topical domains showed evidence that one version of the IAT was more strongly related to self-reported attitudes than the other. We added the target concept error rate variable with interaction terms as additional predictors to the five simultaneous regression models with condition (modified or original), attitudes, knowledge, and interactions predicting IAT score. In all cases, the condition X attitude interaction was no longer significant after adding the error variable to the model (ps range from .11 to .68) suggesting that task recoding is sufficient to account for the attitude-IAT correlation differences between IAT versions.

Summary

The IAT was robustly related to self-reported attitudes and showed negligible relations with evaluative knowledge across 58 attitude domains. Those relations showed minor variation between original and modified IATs, with the modified IAT showing a slightly increased correspondence with self-reported attitudes in a few domains. Replicating Studies 4-6 across a wide variety of content domains, participants were more likely to explicitly evaluate target concepts in the IAT when error feedback was absent and the category labels emphasized one's own evaluation (I like/I dislike) compared to the original IAT.

General Discussion

In seven studies, we investigated the extent to which implicit evaluations measured by the IAT assessed personal attitudinal information versus extra-personal knowledge. Four distinct lines of evidence were consistent with our hypothesis that the IAT is an individual difference measure of evaluation and not a reflection of, or contaminated by, extra-personal knowledge.

First, Studies 2-7 consistently demonstrated a reliable relationship between explicit attitudes and the IAT and little to no relationship between evaluative knowledge and the IAT in contrast to assertions that the IAT is largely a measure of knowledge not attitude (Arkes & Tetlock, in press; Karpinski & Hilton, 2001). Second, procedural modifications suggested by Olson and Fazio (2004) to reduce the influence of extra-personal knowledge on the IAT did not do so (Studies 2-7). Third, those procedural modifications did have an effect on attitude assessment but both IATs contained unique attitudinal variation suggesting that the procedural changes were not simply removing confounding variance (Studies 1-3). Finally, the effect of the IAT modifications appears to have been the instigation of a task recoding confound in the IAT in which target concepts were increasingly likely to be explicitly evaluated instead of categorized based on their intended nominal feature (e.g., food, racial group; Studies 4-7).

Is Knowledge Dead?

Across studies, evaluative knowledge was, at best, only weakly related to IAT scores despite:

(a) evidence for the criterion validity of evaluative knowledge showing it to be valid and variable;

(b) a heterogeneous representation of evaluative knowledge representing perceptions of media, historical, or societal portrayals of target concepts, and estimates of the average person's or most people's liking, warmth or preferences for the target objects;

(c) high power for detecting effects increasing the interpretability of the null relationships (for many of the analyses, the power exceeded .95 providing similar levels of confidence for accepting and rejecting the null hypothesis);

(d) the null relations between IAT effects and knowledge contrasted with replicable moderate to strong relations between IAT effects and self-reported attitudes;

(e) consistent effects across a broad array of content domains. Of the 66 knowledge-IAT comparisons across studies (original IAT only), five (7.5%) showed correspondence that significantly differed from zero ($\alpha = .05$), and three of those five were in the 'wrong' direction, and the overall average relationship was practically zero ($\beta = .004$). The rate of statistical significance, the vanishingly small average effect, and the distribution of significant effects in both tails of the distribution foster

suspicions of Type I error. Notably, if we moved α to .01, a common level for studies with large sample sizes like these, none of the five observed IAT-knowledge relations across studies would be significant.

These data are a challenge to the hypothesis that the IAT is consistently and meaningfully influenced by extra-personal knowledge. And yet, it would be premature to announce the death of the extra-personal confound hypothesis on the basis of this evidence alone.

Potential avenues for identifying a meaningful distinction between personal and extra-personal knowledge in implicit cognition. The present data have the strength of a heterogeneous representation of evaluative knowledge. However, there may be as yet untested forms of extra-personal knowledge that will predict IAT effects. Knowledge was represented here in terms of previous theories that posited that the contaminating knowledge is known and available to the individual but irrelevant to personal evaluation (Arkes & Tetlock, in press; Karpinski & Hilton, 2001; Olson & Fazio, 2004). Olson and Fazio (2004) pointed out that evaluative knowledge could be from any source that does not contribute to one's evaluation. The knowledge items in these studies by no means examined every possible knowledge source. If evidence for other sources of evaluative knowledge can be found to influence IAT scores, then it will be an interesting challenge to resolve why those sources do and the ones examined here do not, especially considering that our measures were concordant with theorists' claims about types of evaluative knowledge presumed to influence IAT effects (Arkes & Tetlock, in press; Karpinski & Hilton, 2001).

It is, in some sense, surprising that the IAT showed little to no relation to evaluative knowledge. Indeed, implicit attitudes are presumed to reflect one's experience in their environments, and evaluative knowledge presumably reflects some of that experience. These data suggest that measuring explicit knowledge is not a good way to capture the experience that is ultimately reflected in implicit evaluation. An additional question, considered below, is whether that experience can sensibly be considered extra-personal if it is not accessible to explicit cognition.

Theoretical conceptions of implicit attitudes suggest that they are introspectively inaccessible reflections of previous experience (Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooler, 2000).

While those theories posit that such experience is attitudinal (Banaji, 2001), it is possible that some of that previous experience impacts implicit measurement but is inert in the everyday behavior of the individual. The present studies suggest that variation in IAT effects that is not shared with self-reported attitudes *is not* explicit evaluative knowledge, but it is not clear what this unique variation *is*. The identity of this unique component of implicit evaluation is of particular interest as it reflects evaluations that are dissociated from self-report (Greenwald & Banaji, 1995). Other lines of research suggest that the variation in implicit evaluation that is non-redundant with self-report also reflects individual differences and has predictive validity (e.g., Nosek, Banaji, & Greenwald, 2002b; Poehlman et al., 2004). Also, theories abound concerning the presumed origins and consequences of implicit attitudes that exist outside of awareness or control (Greenwald & Banaji, 1995; Rudman, 2004; Wilson et al., 2000). Even so, there is still much empirical work to be done to determine whether the unique implicit component of evaluation contains variation that is extra-personal in origin and effect.

So, there are still avenues open to investigation concerning the presence of extra-personal information in implicit evaluation. The present results suggest that a distinction between personal and extra-personal components of implicit evaluation will require conceptual refinement of current theories and a replicable empirical demonstration of relations between IAT effects and non-attitudinal evaluative knowledge. Based on the current evidence, our theoretical position, described next, is that such a distinction may not be meaningful for implicit evaluation.

All of Our Evaluations Implicitly Belong to Us

An attitude is defined as the association between a concept and an evaluation that resides in memory (Fazio, Chen, McDonel, & Sherman, 1982). These associations are presumed to form based on experience, direct and indirect, with attitude objects (Eagly & Chaiken, 1998; Greenwald & Banaji, 1995). Once in memory, concept-evaluation associations may influence perception, attention, judgment, and action providing the basis for the prominence of attitudes in social psychological theory and research.

The presence of concept-evaluation associations in memory does not mean that they will influence cognitive processing in all cases. Drawing on Higgins (1996) distinction between associative

information that is available (stored in memory with the potential to be activated) versus accessible (the activation potential of the associative information), Eagly and Chaiken (1998) point out that an attitudinal response is dynamic and can draw on different aspects of the available concept-evaluation associations (see also Wilson & Hodges, 1992). What associative information is activated and influential will depend on its availability, accessibility, and applicability (Higgins, 1996).

The IAT, like other implicit measures, is thought to measure concept-evaluation associations that have developed from experience through mechanisms such as classical conditioning (Olson & Fazio, 2001). Culturally-bound experience is comprised by nationality, state, city, neighborhood, school, family, birth order, friend, gender, ethnicity, age, social class, spoken language, occupation, and any number of other social categories and contexts. Implicit evaluations are presumed to reflect variations in those experiences. Experience may be culturally-bound or culturally-independent, but that distinction is irrelevant for implicit evaluation. What is important for implicit evaluation is that experience must happen, associations between concepts and evaluations must form, and those associations must be available in memory.

Where we differ from Karpinski and Hilton (2001) and Arkes and Tetlock (in press) is that we argue that endorsement, especially in the context of implicit cognition, is irrelevant for information to be a measure of individual attitude and predict individual behavior (Banaji, Nosek, & Greenwald, in press). Where we differ from Olson and Fazio (2004) is that we argue that any evaluative information, no matter how it was learned, is potentially attitudinal and influential for individual thinking, feeling, and acting. Declaring that some information in our own heads is not *personal* may inappropriately focus attention on the source of the information – where we learned it, rather than the consequences of the information – what we do with it. It is in the presence and consequences of information, not the origins, that ownership is established.

Distinguishing Myself from My Knowledge

The preceding discussion might appear to suggest that humans are slaves to their knowledge and experience, and that knowing something is akin to believing it. Humans do appear to represent and

believe information in a singularly Spinozan process (Gilbert, 1991). But, humans also have the remarkable ability to unbelieve things that they once thought and believed. Distinguishing knowledge that is 'mine' from 'just the stuff that I know' is where explicit cognition has a decided advantage over implicit cognition.

A luxury of our conscious processes is that we get to decide whether we accept or reject the information that bubbles up from memory. Stereotypes about racial, gender, age, or political groups can come to mind and be accepted or rejected. Also, we can invoke higher-order principles for informing on our judgments and actions toward group members, such as "treat others not by the color of their skin but the content of their character." These explicit processes provide opportunities to effortfully correct biases present in the culture or our own mind that may conflict with the ways in which we want to perceive, judge, and act toward others, or attitude objects in general.

Implicit or automatic processes that operate outside of conscious awareness or conscious control afford few such corrective mechanisms. The information available in memory, whatever the source and whether personally accepted or rejected, can influence perception, judgment, and action whenever it becomes actively involved in cognitive processing. Whether certain information becomes influential may be determined by multiple processes such as chronic goals or motivations (Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002; Moskowitz, Gollwitzer, Wasel, & Schaal, 1999) or by the degree to which information is well-learned, situationally-relevant, or immediately accessible (Higgins, 1996).

Avoiding the influence of concept-evaluation associations that we would prefer not to claim as our own requires awareness of their presence, capacity to exert control over their expression, and the knowledge or skill to correct for their influence. This may not be simple. Consider the phenomenon of stereotype or identity threat in which members of stereotyped groups show performance decrements in the stereotyped domain when the relevant stereotype or social identity is activated (Steele & Aronson, 1995; Steele, Spencer, & Aronson, 2002). The impact of the stereotype knowledge need not be chronically accessible, personally endorsed, or even available to conscious awareness in order to have its insidious impact - it need only be activated (Dijksterhuis, Aarts, Bargh, & van Knippenberg, 2000; Dijksterhuis &

van Knippenberg, 1998; Shih, Pittinsky, & Ambady, 1999; Steele et al., 2002). It might be more comfortable to say that those known stereotypes are 'not mine' because they are explicitly disavowed and a threat to self. Nonetheless, those stereotypes are in mind and influential, making them unavoidably, even undesirably, one's own.

The selves that we are and the selves we intend to be are both *us*, and sometimes they do not agree. One could say that humans are large, containing multitudes. Full recognition of this fact raises serious questions for important issues of responsibility, culpability, and intentionality. When should organisms (even human ones) be held responsible for their actions? What role should intentionality play in drawing the line between the responsible agent and the causal, but not responsible, agent? These issues reach far beyond the penultimate paragraph of a paper, but are ones that psychologists, ethicists, and legal analysts must continue to scrutinize.

All concept-evaluation associations that are available in memory have the potential to influence processing, perception, judgment, and action – so, all such associations are attitudinal. Efforts to understand when, why, and how various aspects of those attitudes will have influence should keep social psychologists busy and ensure that attitudes will remain one of the field's most indispensable constructs.

Footnotes

[1] When analyzing simple correlations as Olson and Fazio (2004) did rather than as disattenuated correlations in SEM, the difference in correlation magnitude was reliable ($p = .02$).

[2] For most analyses reported in this paper, we conducted follow-up tests to the structural equation models on each of the measured indicators of evaluative knowledge (Studies 2-6) to show that the reported observations were consistent across the heterogeneous construct. For example, six of the seven simultaneous regressions of each explicit attitude and behavioral intention measures regressed on the original and modified IATs showed that the original IAT captured unique attitudinal variation (lone exception: $p = .07$). Follow-up tests reinforced the generality of the observations reported in the structural models unless otherwise reported.

[3] Note that equality constraints from the models concern unstandardized coefficients and all Figures present standardized coefficients for interpretability. Also, in all studies there was no difference in results if questions 3 and 4 were evaluated before questions 1 and 2.

[4] Another 98 participants did at least one task but did not finish the entire study either because of technical malfunction, distraction, or disinterest. The study non-completers were slightly younger ($M = 25$, $SD = 10.2$), but undifferentiated across gender and ethnicity.

[5] Typically, using exploratory factor analysis is considered a liability because it can increase the incidence of Type I error, and a priori factor identification is much preferred. However, this approach maximizes the potential to show that a knowledge-IAT relation may exist. Note that the exploratory factor identification was only conducted for knowledge. For explicit attitudes, a single a priori defined factor was used in all models.

[6] This is one more domain than was examined by Nosek (2005). We also examined attitudes toward Burger King compared to McDonalds.

[7] One factor that may influence the relationship between attitudes and knowledge with other variables is the amount of variability in those ratings. If there were consensus in perceptions of the average person's attitude, then there would be no meaningful variability that could relate to other variables. Studies 4, 5 and 6 showed that there was variability in perceived others' evaluations and it predicted criterion variables in those three domains. In Study 7, while (unsurprisingly) the variability in evaluative knowledge tended to be less than individual attitude responses (attitude variability across domains, $SD = 2.8$, knowledge variability across domains, $SD = 2.1$) there existed substantial variability in perceived cultural attitudes in all cases (see Table 1; SD range = 1.3 - 3.0). Remarkably, in 12 of the topical domains (21%), variability in the perceived cultural evaluation actually exceeded that of the self-reported attitudes. In other words, people's perceptions of what the *average person* preferred actually varied more than *individual* preferences themselves directly opposing the trends that one would expect if perceptions of group averages conformed, even a little, to the expectations of the central limit theorem. While perceptions of others' evaluations are less variable on average than individual attitudes, it is clear that they are themselves quite variable, and the observed effects are no different for domains in which individual attitudes are more or less variable than evaluative knowledge.

References

Arkes, H. R., & Tetlock, P. E. (in press). Attributions of implicit prejudice, or "would Jesse Jackson 'fail' the Implicit Association Test?". *Psychological Inquiry*.

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, J. S. Nairne, I. Neath, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117-150). Washington D.C.: American Psychological Association.

Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (in press). No place for nostalgia in science: A response to Arkes & Tetlock. *Psychological Inquiry*.

Bargh, J. A. (1996). Automaticity in social psychology. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 169-183). New York: Guilford.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Thousand Oaks, CA: Sage.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. (2002). The regulation of explicit and implicit racial bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology, 82*(5), 835-848.

Dijksterhuis, A., Aarts, H., Bargh, J. A., & van Knippenberg, A. (2000). On the relation between associative strength and automatic behavior. *Journal of Experimental Social Psychology, 36*, 531-544.

Dijksterhuis A., & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of Trivial Pursuit. *Journal of Personality and Social Psychology, 74*, 865-877.

Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 269-322). Boston, MA: McGraw-Hill.

Fazio, R. H., Chen, J., McDonel, E. C., & Sherman, S. J. (1982). Attitude accessibility, attitude-behavior consistency, and the strength of the object-evaluation association. *Journal of Experimental Social Psychology, 18*, 339-357.

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

Fiske, A. P., Kitayama, S., Markus, H. R., & Nisbett, R. E. (1998). The cultural matrix of social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 915-981). Boston, MA: McGraw-Hill.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46*, 107-119.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4-27.

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., 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.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133-168). New York: Guilford Press.

Hofman, W., Gawronski, B., Gschwendner, T., Le, H. (2004). *A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures*. Unpublished manuscript, University of Trier, Germany.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*, 117-144.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 774-788.

MacCullum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130-149.

Markus, H. R., Kitayama, S., & Heiman, R. J. (1996). Culture and "basic" psychological principles. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 857-913), New York: Guilford Press.

McArdle, J. J., & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. Marcoulides & R. Schumacker (Eds.), *Advanced Structural Equation Modeling Techniques.* Hillsdale, NJ: Erlbaum. 89-124.

McArdle, J. J., Johnson, R. C., Hishinuma, E. S., Miyamoto, R. H., & Andrade, N. N. (2001). Structural equation modeling of group differences in CES-D ratings of Native Hawaiian and non-Hawaiian high school students. *Journal of Adolescent Research, 16*, 108-149.

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika, 29*, 177-185.

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77*(1), 167-184.

Nosek, B. A. (2005). *Multi-level modeling of the relationship between implicit and explicit attitudes.* Unpublished manuscript, University of Virginia.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002a). Harvesting intergroup implicit attitudes and beliefs from a demonstration Web site. *Group Dynamics, 6*(1), 101-115.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002b). Math = me, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology, 83*(1), 44-59.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (in press). Understanding and using the Implicit Association Test: II. Methodological issues. *Personality and Social Psychology Bulletin.*

Nosek, B. A., & Smyth, F. L. (2004). *A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs*. Unpublished manuscript, University of Virginia.

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

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*(5), pp. 653-667.

Poehlman, T. A., Uhlmann, E., Greenwald, A. G., & Banaji, M. R. (2004). *Measuring and using the Implicit Association Test: III. Meta-analysis of predictive validity*. Unpublished manuscript, Yale University.

Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science, 13*(2), 79-83.

Saucier, D. A., & Miller, C. T. (2003). The persuasiveness of racial arguments as a subtle measure of racism. *Personality and Social Psychology Bulletin, 29*, 1303-1315.

Sears, D. O. (in press). One survey researcher's perspective on implicit prejudice. *Psychological Inquiry*.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80-83.

Steele, C. M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797-811.

Steele, C. M., Spencer, S. J. & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.) *Advances in experimental social psychology* (Vol. 34, pp. 379-440). San Diego, CA: Academic Press.

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

Steiger, J. H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduck and Glaser. *Structural Equation Modeling, 7*, 149-162.

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of factors*. Paper presented at the annual spring meeting of the Psychometric Society, Iowa City, IA.

Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of The Vectors of Mind*. Chicago: The University of Chicago Press.

Wegner, D.M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.

Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D.T. Gilbert, S.T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 446-496). Boston, MA: McGraw-Hill.

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: Harvard 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.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*(1), 101-126.

Appendix

Factor	Observed Variable	Item(s)	Scale	Mean	Std
Study 1					
EA	GBSD: Difference score of mean semantic differential ratings (Gore - Bush)	George Bush [Al Gore] is…	mean of 5 ratings (1-7: unattractive-attractive, awful-nice, unpleasant-pleasant, wise-foolish, good-bad)	0.2	2.1
EA	GBLIKE: Difference score of scale ratings (Gore - Bush)	How much do you like George Bush [Al Gore]?	1-7: not at all-very much	0.4	2.8
EA	GBFAV: Difference score of thermometer ratings (Gore - Bush)	How favorable do you feel towards George Bush [Al Gore]?	0-100: extremely unfavorable-extremely favorable	5.9	44.4
EA	ELECT: Semantic differential rating	If an election involving Bush and Gore for president were held today, for whom would you vote?	1-7: Bush-Gore	4.2	2.4
EA	ESD: Semantic differential ratings	Who is more intelligent, likeable, qualified and has stronger character?	mean of 4 ratings (1-7: Bush-Gore)	4.2	1.5
Study 2					
EA	WARM: Difference score of thermometer ratings (White - Black)	How warmly do you feel toward Black [White] Americans?	0-10: extremely cold-extremely warm	0.6	1.6
EA	LIKE: Difference score of scale ratings (White - Black)	How much do you like Black [White] Americans?	1-7: not at all-very much	0.3	1.0
EK	HIST: Difference score of scale ratings (White - Black)	Historically, how favorably or unfavorably has American society been for Black [White] Americans?	1-7: very unfavorable-very favorable	3.7	1.9
EK	CFAV: Difference score of thermometer ratings (White - Black)	How warmly does American society feel toward Black [White] Americans?	0-10: extremely cold-extremely warm	2.9	2.0
EK	CLIKE: Difference score of scale ratings (White - Black)	How favorably or unfavorably does American society portray Black [White] Americans?	1-7: very unfavorable-very favorable	2.4	1.6
Study 3					
EA	FAV: Difference score of thermometer ratings (Peanuts - Shellfish)	How favorable do you feel toward peanuts [shellfish]?	0-100: extremely unfavorable-extremely favorable	15.7	47.4
EA	EAT: Difference score of scale ratings (Peanuts - Shellfish)	How much do you like to eat peanuts [shellfish]?	1-6: very little-very much	0.5	2.5
EA	LIKE: Difference score of scale ratings (Peanuts - Shellfish)	How much do you like peanuts [shellfish]?	0-10: not at all-a lot	1.3	4.6
EK	CFAV: Difference score of scale ratings (Peanuts - Shellfish)	How favorable is American culture toward peanuts [shellfish]?	1-6: very unfavorable-very favorable	0.9	1.5
EK	CLIKE: Difference score of scale ratings (Peanuts - Shellfish)	How much does the average person like peanuts [shellfish]?	1-6: very little-very much	1.0	1.2
Study 4					
EA	CHAR: Difference score of mean semantic differential ratings (Bush - Kerry)	Who is more intelligent, likeable, qualified and has stronger character?	mean of 4-items (1-7: John Kerry-George Bush)	-0.9	2.4
EA	LIKE: Difference score of scale ratings (Bush - Kerry)	How much do you like George Bush [John Kerry]?	1-7: not at all-very much	-1.2	2.9
EA	TEMP: Difference score of thermometer ratings (Bush - Kerry)	How favorable do you feel towards George Bush [John Kerry]?	0-100: very unfavorable-very favorable	-23.7	51.5

Appendix Continued

EA	VOTE: Semantic differential item	If an election involving George Bush and John Kerry for president were held today, for whom would you vote?	1-7: John Kerry-George Bush	2.8	2.4
EK	AVG: Semantic differential item	Does the average person prefer George Bush or John Kerry?	1-7: John Kerry-George Bush	3.8	1.6
EK	MOST: Semantic differential item	Do most people prefer George Bush or John Kerry?	1-7: John Kerry-George Bush	3.8	1.5
EK	LIKE: Difference score of scale ratings (Bush - Kerry)	How much does the average person like or dislike George Bush [John Kerry]?	1-6: strongly dislikes-strongly likes	-0.4	1.6
EK	WARM: Difference score of scale ratings (Bush - Kerry)	How warm or cold is society to George Bush [John Kerry]?	1-6: very cold-very warm	-0.4	1.7
EK	HIST: Difference score of scale ratings (Bush - Kerry)	Historically, how favorable or unfavorable has American society been towards George Bush [John Kerry]?	1-6: very unfavorable-very favorable	0.5	1.7
EK	PORT: Difference score of scale ratings (Bush - Kerry)	How favorably or unfavorably does American society portray George Bush [John Kerry]?	1-6: very unfavorably-very favorably	-0.2	1.9
Study 5					
EA	WARM: Difference score of thermometer ratings (Black - White)	How warmly do you feel toward Black [White] Americans?	0-10: extremely cold-extremely warm	-0.2	1.8
EA	LIKE: Difference score of scale ratings (Black - White)	How much do you like Black [White] Americans?	1-7: not at all-very much	-0.1	1.1
EK1	LIKE: Difference score of scale ratings (Black - White)	How much does the average person like or dislike Black [White] Americans?	1-6: strongly dislikes-strongly likes	-0.8	1.2
EK1	AVG: Semantic differential item	Does the average person prefer Black Americans or White Americans?	1-7: White Americans-Black Americans	2.8	1.2
EK1	WARM: Difference score of scale ratings (Black - White)	How warm or cold is society to Black [White] Americans?	1-6: very cold-very warm	-2.1	1.5
EK2	HIST: Difference score of scale ratings (Black - White)	Historically, how favorable or unfavorable has American society been toward Black [White] Americans?	1-6: very unfavorable-very favorable	-3.9	1.5
EK2	PORT: Difference score of scale ratings (Black - White)	How favorably or unfavorably does American society portray Black [White] Americans?	1-6: very unfavorably-very favorably	-2.4	1.7
EK2	MOST: Semantic differential item	Do most people prefer Black Americans or White Americans?	1-7: White Americans-Black Americans	2.9	1.2
Study 6					
EA	ATT: Difference score of mean semantic differential ratings (candy bars - apples)	Candy bars [Apples] are…	5 ratings (1-7: ugly-beautiful, horrible-wonderful, disgusting-tasty, bad-good, unpleasant-pleasant)	-0.7	1.4
EA	TEMP: Difference score of thermometer ratings (candy bars - apples)	How favorable do you feel toward candy bars [apples]?	0-100: very unfavorable-very favorable	-9.1	32.3
EK1	LIKE: Difference score of scale ratings (candy bars - apples)	How much does the average person like or dislike candy bars [apples]?	1-6: strongly dislikes-strongly likes	0.7	1.0
EK1	WARM: Difference score of scale ratings (candy bars - apples)	How warm or cold is society to candy bars [apples]?	1-6: very cold-very warm	-0.4	1.6
EK1	AVG: Semantic differential item	Does the average person prefer candy bars or apples?	1-7: candy bars-apples	4.4	1.5
EK1	MOST: Semantic differential item	Do most people prefer candy bars or apples?	1-7: candy bars-apples	4.3	1.5
EK2	HIST: Difference score of scale ratings (candy bars - apples)	Historically, how favorable or unfavorable has American society been toward candy bars [apples]?	1-6: very unfavorable-very favorable	-0.6	1.5
EK2	PORT: Difference score of scale ratings (candy bars - apples)	How favorably or unfavorably does American society portray candy bars [apples]?	1-6: very unfavorably-very favorably	-1.0	1.9

Figure Captions

Figure 1. Structural equation model of explicit racial attitudes (EA) and evaluative knowledge (EK) predicting original (oIAT) and modified (mIAT) IATs (Study 2, Model 3 from Table 1).

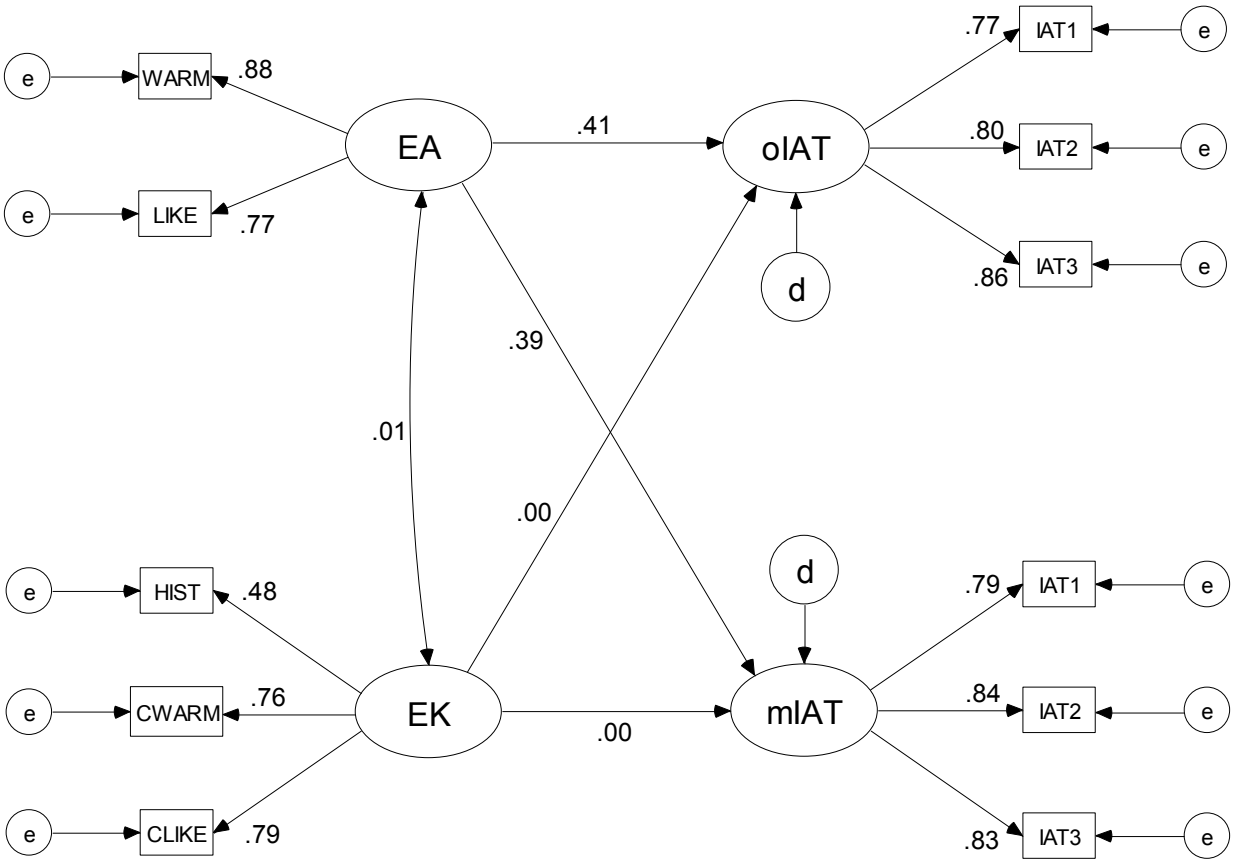
Figure 2. Structural equation model of explicit food attitudes (EA) and evaluative knowledge (EK) predicting original (oIAT) and modified (mIAT) IATs (Study 3, Model 3 from Table 1).

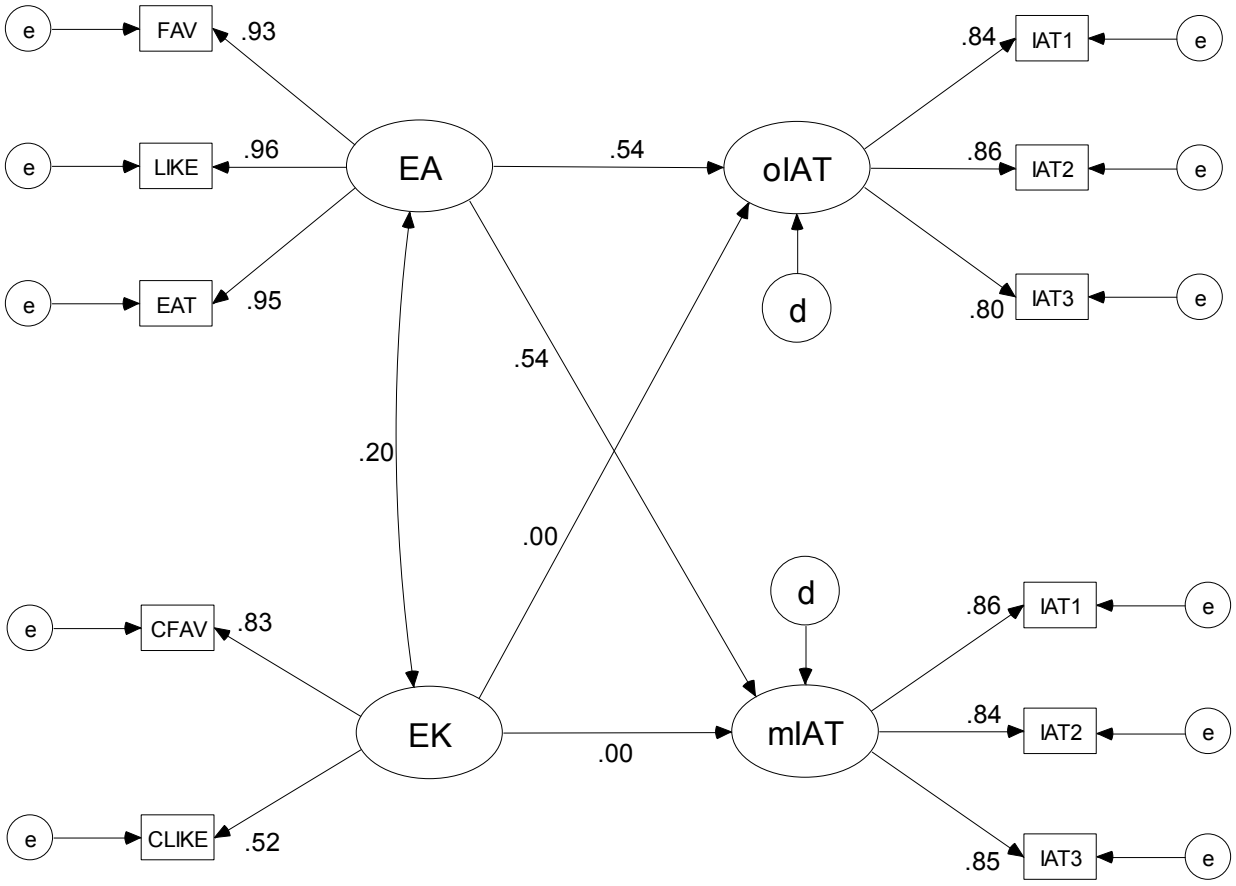
Figure 3. Two-group (original IAT top and modified IAT bottom) structural equation model of explicit political attitudes (EA) and evaluative knowledge (EK) predicting IAT scores (Study 4, Model 1 from Table 2).

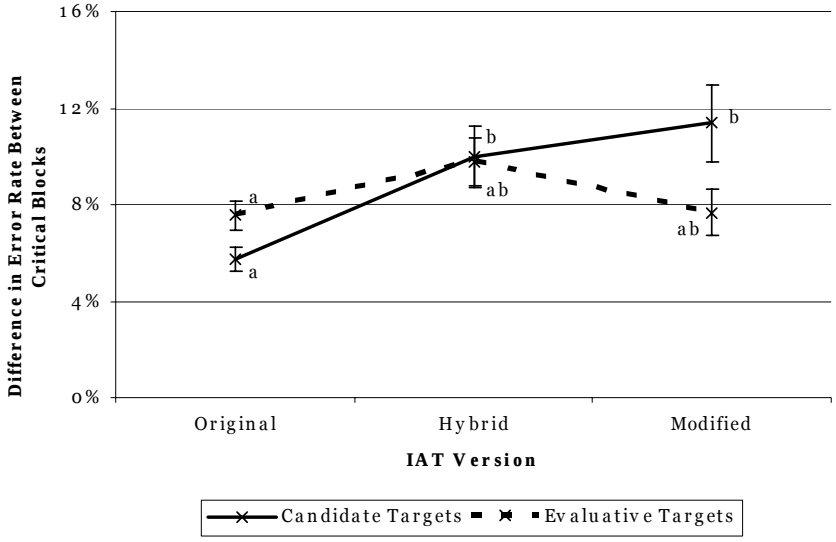
Figure 4. Mean absolute difference in errors between critical combined blocks for Bush-Kerry IAT (Study 4), Black-White IAT (Study 5), and Candy Bar-Apple IAT (Study 6) across original, hybrid, and modified IAT task procedures. Error bars reflect standard error of the mean. Means with different letters within exemplar sets (target concepts or evaluative) are significantly different from one another ($p < .05$).

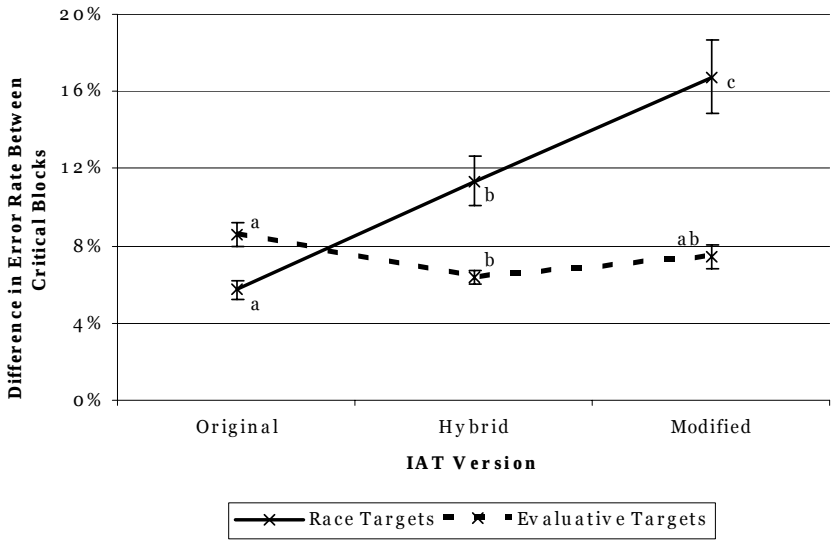
Figure 5. Two-group (original IAT top and modified IAT bottom) structural equation model of explicit racial attitudes (EA) and evaluative knowledge (EK) predicting IAT scores (Study 5, Model 3 from Table 2).

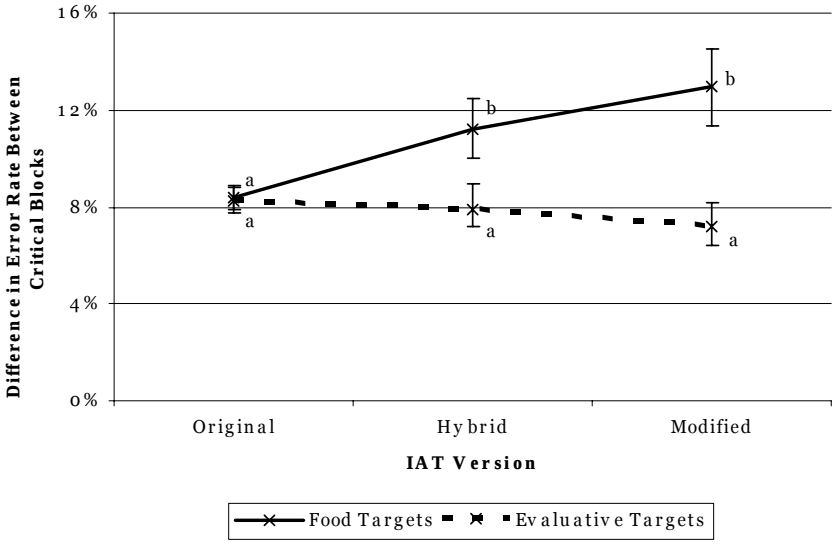
Figure 6. Two-group (original IAT top and modified IAT bottom) structural equation model of explicit food attitudes (EA) and evaluative knowledge (EK) predicting IAT scores (Study 6, Model 2 from Table 2).

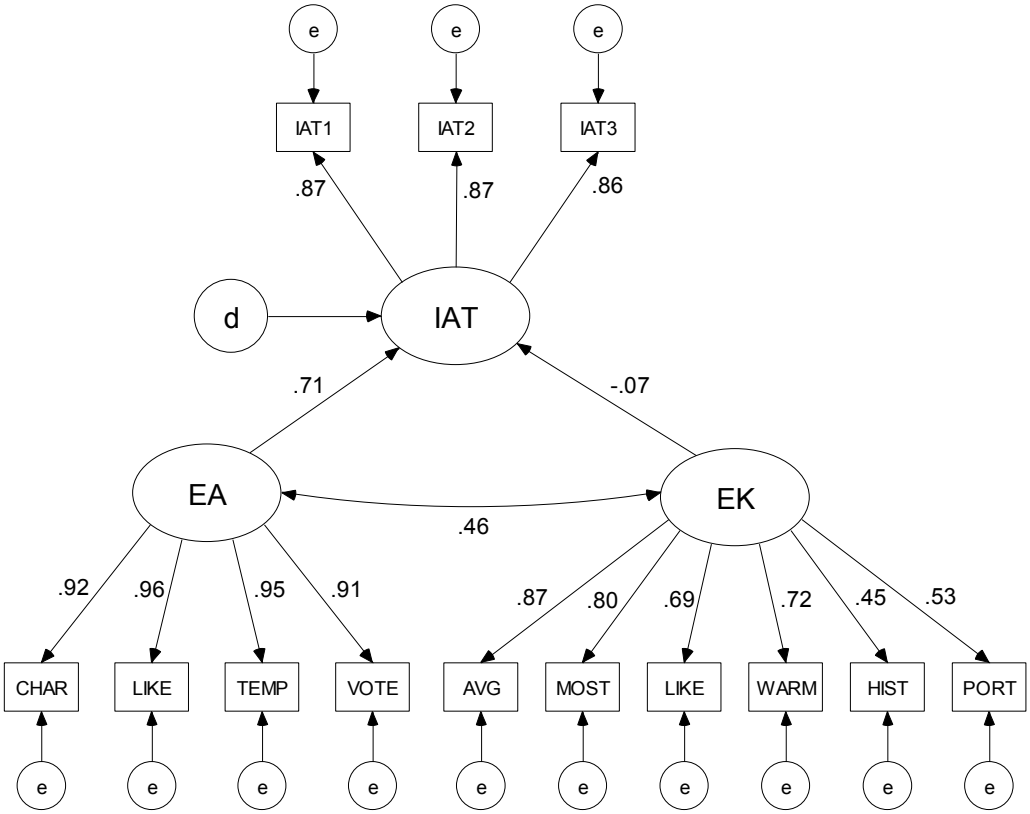

e
WARM
.88
e
LIKE
.77
EA
.41
oIAT
.77
IAT1
e
.80
IAT2
e
IAT3
.86
e
d
.39
.01
.00
e
HIST
.48
e
CWARM
.76
e
CLIKE
.79
EK
.00
d
mIAT
.79
IAT1
e
.84
IAT2
e
.83
IAT3
e

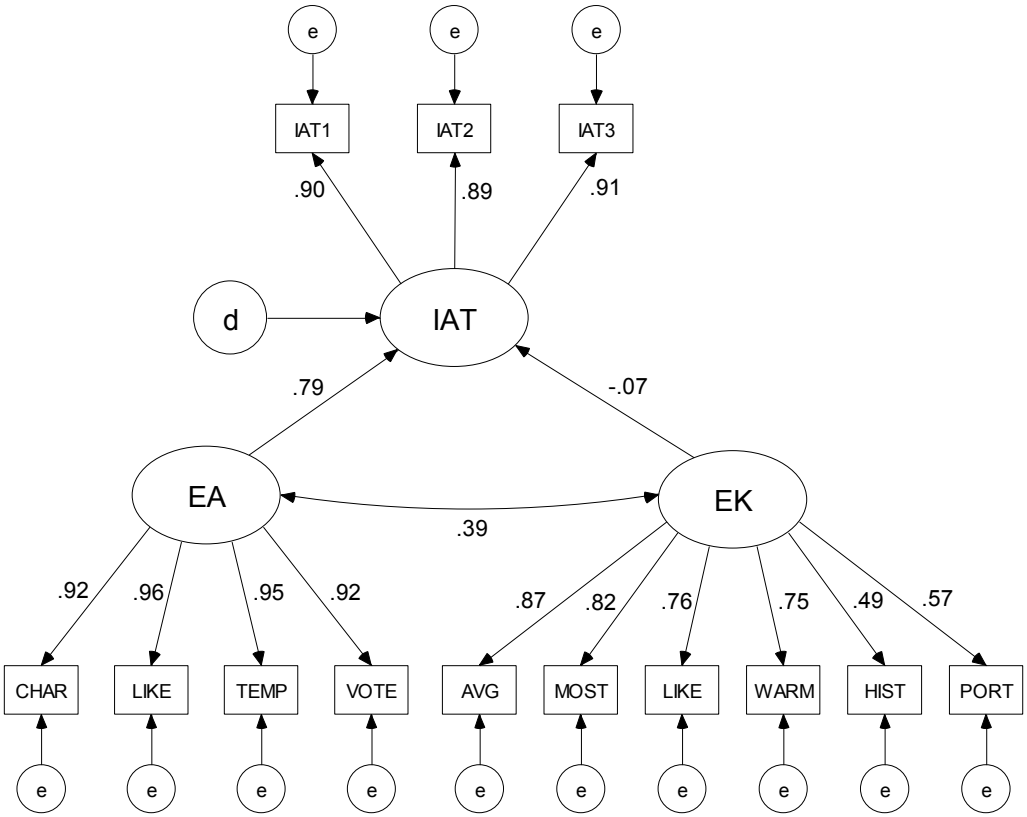

e
FAV
.93
e
LIKE
.96
EA
.54
oIAT
.84
IAT1
e
.86
IAT2
e
e
EAT
.95
d
.80
IAT3
e
.54
.20
.00
d
e
CFAV
.83
.86
IAT1
e
EK
.00
mIAT
.84
IAT2
e
e
CLIKE
.52
.85
IAT3
e

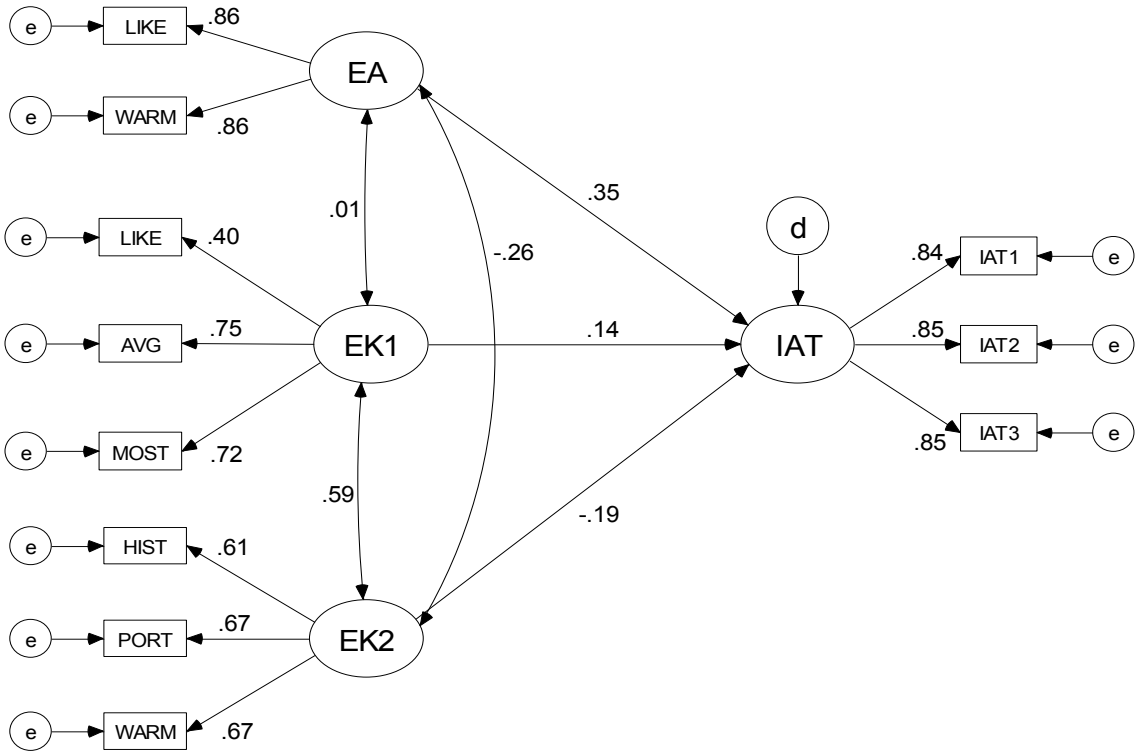

16%
12%
8%
4%
0%
Difference in Error Rate Between Critical Blocks
a
a
b
ab
ab
b
Original
Hybrid
Modified
IAT Version
Candidate Targets
Evaluative Targets

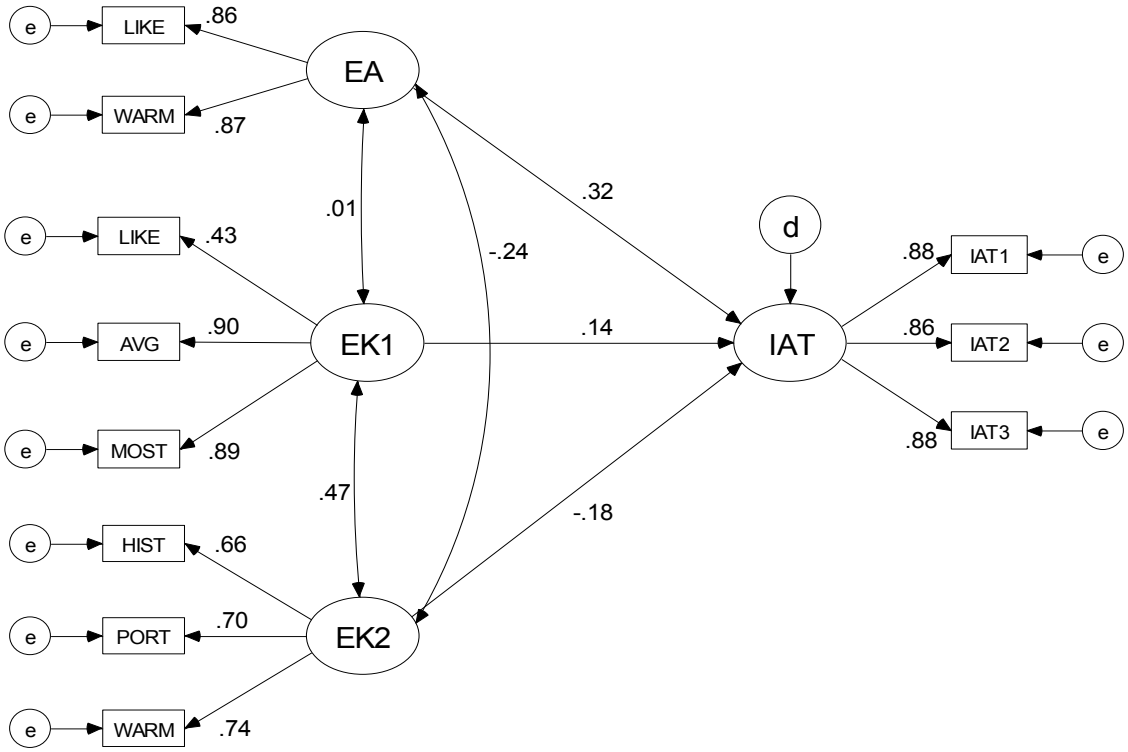

20%
16%
12%
8%
4%
0%
Difference in Error Rate Between Critical Blocks
a
a
b
b
ab
c
Original
Hybrid
Modified
IAT Version
Race Targets
Evaluative Targets

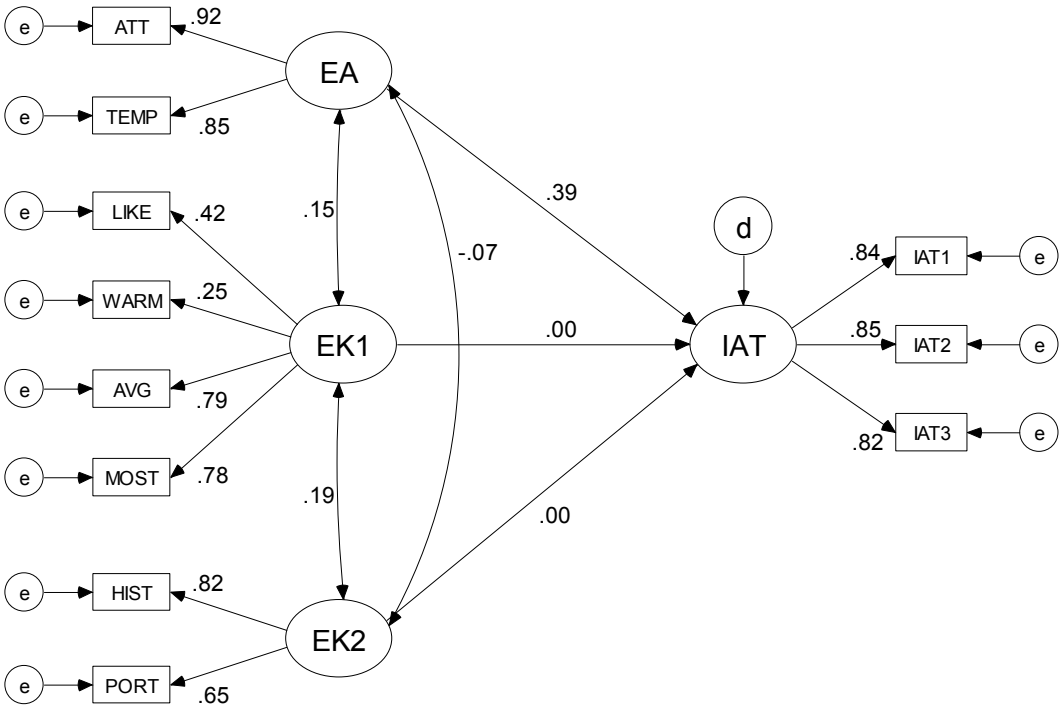

16%
12%
8%
4%
0%
Difference in Error Rate Between Critical Blocks
a
a
b
a
a
b
Original
Hybrid
Modified
IAT Version
Food Targets
Evaluative Targets


e
e
e
IAT1
IAT2
IAT3
.87
.87
.86
d
IAT
.71
-.07
EA
EK
.46
.92
.96
.95
.91
.87
.80
.69
.72
.45
.53
CHAR
LIKE
TEMP
VOTE
AVG
MOST
LIKE
WARM
HIST
PORT
e
e
e
e
e
e
e
e
e
e


e
e
e
IAT1
IAT2
IAT3
.90
.89
.91
d
IAT
.79
-.07
EA
EK
.39
.92
.96
.95
.92
.87
.82
.76
.75
.49
.57
CHAR
LIKE
TEMP
VOTE
AVG
MOST
LIKE
WARM
HIST
PORT
e
e
e
e
e
e
e
e
e
e


e
LIKE
.86
EA
e
WARM
.86
.01
.35
d
e
LIKE
.40
-.26
.84
IAT1
e
e
AVG
.75
EK1
.14
IAT
.85
IAT2
e
e
MOST
.72
.85
IAT3
e
.59
-.19
e
HIST
.61
e
PORT
.67
EK2
e
WARM
.67


e
LIKE
.86
EA
e
WARM
.87
.01
.32
d
e
LIKE
.43
-.24
.88
IAT1
e
e
AVG
.90
EK1
.14
IAT
.86
IAT2
e
e
MOST
.89
.88
IAT3
e
.47
-.18
e
HIST
.66
e
PORT
.70
EK2
e
WARM
.74


e
ATT
.92
EA
e
TEMP
.85
e
LIKE
.42
.15
-.07
.39
d
.84
IAT1
e
e
WARM
.25
EK1
.00
IAT
.85
IAT2
e
e
AVG
.79
.82
IAT3
e
e
MOST
.78
.19
.00
e
HIST
.82
EK2
e
PORT
.65

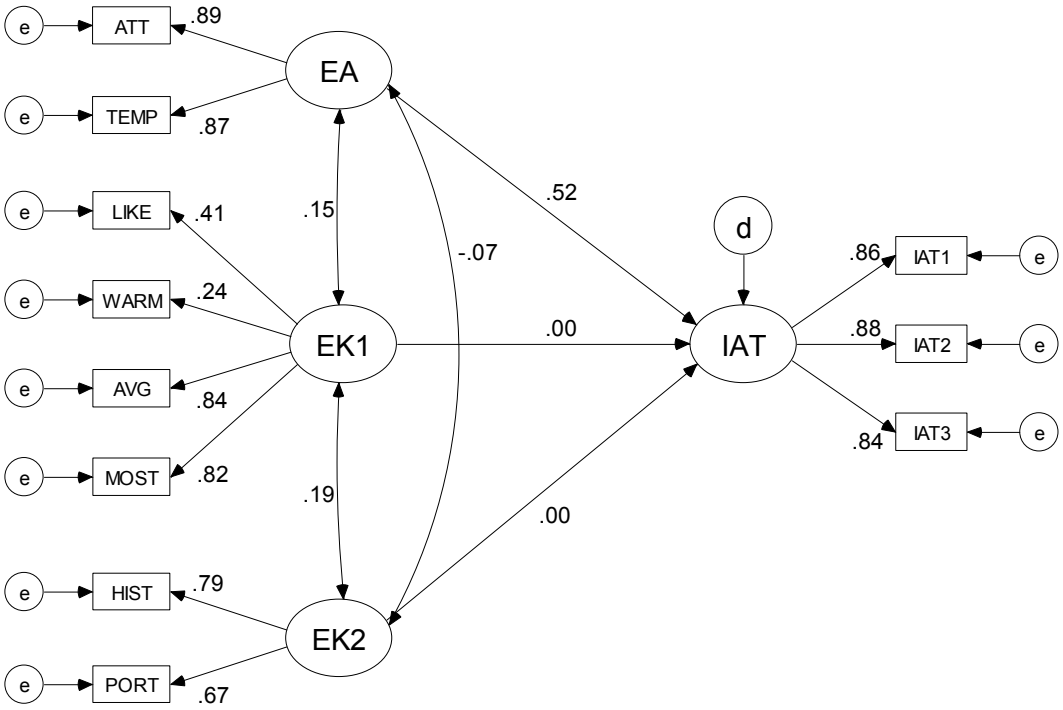

e
ATT
.89
EA
e
TEMP
.87
e
LIKE
.41
.15
-.07
.52
d
.86
IAT1
e
e
WARM
.24
EK1
.00
IAT
.88
IAT2
e
e
AVG
.84
.84
IAT3
e
e
MOST
.82
.19
.00
e
HIST
.79
EK2
e
PORT
.67

Table 1. Hypothesis Tests Relating Explicit Attitudes, Evaluative Knowledge, and the IAT for Studies 2 and 3 Presented in a Series of Structural Model Fit Comparisons.

Model	χ^2	df	$\Delta\chi^2$/df	ε_a	95% CI ε_a of Δ
Study 2 - Racial Attitudes					
0) All parameters free	62	39		.065	
1) EK->mIAT, EK->oIAT constrained equal	62	40	.2 / 1	.063	.000, .199
2) EK->mIAT, EK->oIAT fixed to 0	62	41	.1 / 1	.061	.000, .183
3) M2 + EA->mIAT, EA->oIAT constrained equal	62	42	0 / 1	.061	.000, .000
4) M3 + EA->IAT fixed to 0	94	43	31.6 / 1	.092	.308, .638
Study 3 - Food Attitudes					
0) All parameters free	72	39		.061	
1) EK->mIAT, EK->oIAT constrained equal	72	40	0 / 1	.059	.000, .000
2) EK->mIAT, EK->oIAT fixed to 0	73	41	.8 / 1	.058	.000, .188
3) M2 + EA->mIAT, EA->oIAT constrained equal	73	42	0 / 1	.056	.000, .000
4) M3 + EA->IAT fixed to 0	189	43	115.6 / 1	.121	.578, .836

Note: All comparisons were to prior model. EA=Explicit Attitude; EK=Evaluative Knowledge (numbered if represented as multiple correlated factors); IAT = Implicit Association Test; ε_a = root mean square error of approximation; CI = confidence interval; mIAT = modified IAT; oIAT = original IAT.

Table 2. Hypothesis Tests Relating Explicit Attitudes, Evaluative Knowledge and the IAT for Studies 4, 5, and 6 Presented in a Series of Structural Model Fit Comparisons.

Model	χ^2	df	$\Delta\chi^2$/df	ε_a	95% CI ε_a of Δ
Study 4 - Political Attitudes					
0) All parameters free	477	124		.062	
0.1) Factor loadings constrained equal across groups	493	133	16.5 / 9	.061	.000, .063
0.2) EA<->EK cequal across groups	494	134	.4 / 1	.061	.000, .095
1) M0.2 + EK->IAT equal across groups	494	135	.5 / 1	.060	.000, .098
2) M1 + EK->IAT fixed to 0	498	136	3.9 / 1	.060	.000, .146
3) M1 + EA->IAT equal across groups	502	136	7.5 / 1[a]	.061	.029, .174[a]
4o) M1 + EA->IAT fixed to 0 for Original IAT only	661	136	166.8 / 1[a]	.073	.405, .550[a]
4m) M1 + EA->IAT fixed to 0 for Modified IAT only	750	136	255.3 / 1[a]	.079	.518, .663[a]
Study 5 - Racial Attitudes					
0) All parameters free	118	76		.036	
0.1) Factor loadings constrained equal across groups	125	83	6.8 / 7	.034	.000, .065
0.2) M0.1 + EA<->EK1, EA<->EK2 equal across groups	126	86	1.7 / 3	.033	.000, .076
1) M0.2 + EK1->IAT, EK2->IAT equal across groups	127	88	1.1 / 2	.032	.000, .092
2) M1 + EK1->IAT, EK2->IAT fixed to 0	133	90	6.0 / 2	.034	.000, .146
3) M1 + EA->IAT equal across groups	128	89	.4 / 1[a]	.032	.000, .125[a]
4) M3 + EA->IAT fixed to 0	158	90	30.6 / 1	.042	.173, .363
Study 6 - Food Attitudes					
0) All parameters free	259	76		.056	
0.1) Factor loadings equal across groups	261	83	.9 / 7	.053	.000, .000
0.2) EA<->EK1, EA<->EK2, EK1<->EK2 equal across groups	263	86	1.9 / 3	.052	.000, .059
1) M0.2 + EK1->IAT, EK2->IAT equal across groups	264	88	1.3 / 2	.051	.000, .072
2) M1 + EK1->IAT, EK2->IAT, fixed to 0	266	90	1.7 / 2	.051	.000, .077
3) M2 + EA->IAT equal across groups	270	91	4.5 / 1	.051	.000, .148
4o) M2 + EA->IAT fixed to 0 for Original IAT only	309	91	43.3 / 1[a]	.056	.168, .310[a]
4m) M2 + EA->IAT fixed to 0 for Modified IAT only	346	91	80.5 / 1[a]	.061	.255, .397[a]

Note: [a] compared to Model 1 in top two panels and to Model 2 in bottom panel, otherwise all comparisons were to prior model. EA=Explicit Attitude; EK=Evaluative Knowledge (numbered if represented as multiple correlated factors); IAT = Implicit Association Test; ε_a = root mean square error of approximation; CI = confidence interval.

Table 3. Descriptive Statistics, Regression Beta-Weights of Attitudes and Knowledge Predicting Original and Modified IATs, and Error Means and Effect Sizes of Differences For Task Recoding across 58 domains (Study 7).

			IAT		Attitude		Knowledge		Regressions Summary (Betas)				Task Recoding		
									Original		Modified		Concept Error Differences		
Implicitly Liked	Implicitly Disliked	N	Mean	STD	Mean	STD	Mean	STD	EA->IAT	EK->IAT	EA->IAT	EK->IAT	Original	Modified	d
Democrats	Republicans	176	.22	(.54)	1.7	(4.1)	.0	(1.7)	.73*	-.04	.71*	-.07	.07	.14	.46*
Gore	Bush	189	.09	(.53)	1.2	(4.3)	-.3	(2.5)	.72*	.04	.70*	-.03	.06	.10	.26
Creationism	Evolution	198	.10	(.50)	-1.6	(4.9)	-.5	(3.0)	.70*	-.09	.63*	-.17	.05	.15	.62*
Teen Pop	Jazz	218	.09	(.49)	-1.2	(3.0)	-.6	(2.1)	.67*	.11	.57*	.02	.05	.12	.50*
Liberals	Conservatives	178	.27	(.56)	2.0	(4.1)	-.7	(2.2)	.66*	.00	.68*	.10	.08	.11	.22
Gun Control	Gun Rights	170	.07	(.46)	2.6	(4.1)	.2	(2.2)	.64*	.16	.53*	-.08	.08	.13	.36*
Pro-Choice	Pro-Life	208	.08	(.51)	1.3	(4.9)	-.8	(2.3)	.63*	.04	.44*	-.17	.10	.14	.26
Feminism	Traditional Values	204	.19	(.49)	.4	(3.0)	-1.9	(2.0)	.58*	-.03	.53*	.07	.08	.12	.33*
Religion	Atheism	223	.42	(.49)	1.3	(4.4)	2.3	(2.4)	.57*	-.06	.36*	-.08	.09	.13	.29*
Classical	Hip Hop	240	.31	(.48)	1.0	(3.5)	-1.1	(2.2)	.56*	-.07	.52*	-.06	.05	.09	.28*
American Places	Foreign Places	209	.34	(.44)	.3	(2.2)	.9	(2.4)	.52*	.03	.23*	.01	.06	.12	.44*
David Letterman	Jay Leno	162	.00	(.41)	.3	(2.6)	-.6	(1.7)	.51*	-.18	.35*	.03	.07	.09	.21
Conforming	Rebellious	205	.35	(.49)	-.5	(3.3)	2.2	(2.7)	.49*	.12	.34*	.09	.11	.13	.21
Cats	Dogs	251	.15	(.45)	-.8	(3.2)	-1.7	(1.8)	.47*	-.12	.50*	-.02	.04	.07	.28*
Coke	Pepsi	209	.13	(.42)	.7	(3.0)	.1	(1.7)	.46*	.12	.46*	-.07	.04	.09	.36*
Summer	Winter	200	.44	(.51)	2.2	(3.4)	2.9	(2.3)	.46*	.03	.37*	.00	.06	.10	.29*
Abstaining	Drinking	222	.28	(.50)	.8	(3.7)	-1.9	(2.2)	.43*	-.06	.39*	-.07	.09	.15	.37*
Vegetables	Meat	232	.29	(.45)	.8	(3.3)	-1.4	(1.9)	.43*	-.02	.53*	-.08	.04	.08	.35*
Social Programs	Tax Reductions	166	.29	(.40)	.9	(3.0)	-1.9	(2.3)	.42*	-.11	.43*	-.10	.08	.12	.30
Leaders	Helpers	205	.06	(.50)	-1.0	(2.0)	-1.2	(2.3)	.41*	.11	.31*	.06	.05	.09	.37*
Books	Television	219	.09	(.45)	1.4	(3.0)	-2.3	(2.1)	.40*	.09	.55*	.07	.06	.08	.23
Microsoft	Apple	183	.01	(.47)	.3	(3.1)	.0	(2.2)	.40*	-.08	.47*	.01	.05	.09	.27
Northerners	Southerners	216	.13	(.52)	.3	(2.4)	.0	(1.9)	.37*	.10	.31*	.04	.05	.08	.26
Relaxing	Exercising	205	.05	(.53)	1.2	(2.6)	2.1	(2.1)	.37*	-.06	.35*	-.01	.06	.12	.41*
Nerds	Jocks	266	.14	(.44)	.7	(2.7)	-2.7	(2.5)	.36*	.16	.29*	-.07	.06	.08	.25*
Yankees	Diamondbacks	171	.17	(.40)	.0	(2.5)	-.2	(2.0)	.35*	.09	.47*	-.12	.05	.12	.40*
Education	Defense	201	.59	(.40)	2.8	(2.3)	.5	(2.1)	.34*	.08	.10	-.01	.05	.09	.36*
Straight People	Gay People	226	.36	(.41)	1.5	(2.6)	3.4	(2.2)	.34*	.04	.34*	-.08	.05	.10	.45*
Public	Private	158	.24	(.48)	-1.3	(2.2)	-1.0	(1.9)	.33*	.17	.24*	.10	.09	.11	.13
Tea	Coffee	204	.08	(.44)	.5	(3.1)	-1.5	(1.8)	.33*	.03	.61*	-.19*	.04	.10	.41*
Christian	Jewish	231	.36	(.43)	.0	(2.0)	.9	(1.8)	.32*	-.18*	.21	-.11	.06	.08	.26*
American	Canadian	215	.26	(.40)	-.1	(2.3)	-1.0	(2.3)	.31*	.05	.28*	.24*	.04	.08	.29*
Married	Single	238	.23	(.50)	1.4	(2.9)	1.8	(2.3)	.31*	-.08	.39*	-.01	.06	.11	.51*
Simple	Difficult	212	.65	(.46)	1.4	(2.8)	2.7	(2.5)	.30*	.05	-.08	.12	.09	.15	.42*
Euro Americans	African Americans	246	.21	(.43)	.2	(1.9)	1.3	(2.1)	.29*	.11	.31*	-.05	.05	.11	.41*
California	New York	202	.15	(.45)	.4	(2.4)	.2	(2.0)	.27*	.07	.32*	-.13	.04	.07	.27
Jews	Muslims	219	.31	(.37)	.9	(1.8)	.8	(1.6)	.26*	.07	.07	.13	.06	.08	.21
Management	Labor	183	.03	(.42)	-.6	(2.5)	-1.4	(2.4)	.26*	-.14	.10	.09	.07	.10	.19
Approaching	Avoiding	164	.76	(.46)	2.3	(2.8)	.9	(2.3)	.25*	-.02	.08	.07	.12	.21	.41*
Meg Ryan	Julia Roberts	223	.02	(.42)	.5	(1.8)	.0	(1.6)	.25*	.20*	.22	.30*	.04	.11	.41*
Tom Cruise	Denzel Washington	217	.13	(.37)	-.9	(1.6)	-.5	(1.3)	.25*	-.12	.18	.16	.05	.08	.26
Thin People	Fat People	240	.34	(.42)	.9	(2.2)	2.8	(2.6)	.24*	.06	-.02	.19	.05	.10	.38*
USA	Japan	230	.38	(.41)	.8	(2.5)	.6	(2.2)	.24*	.08	.22*	-.04	.05	.07	.25
Imprisonment	Capital Punishment	198	.20	(.39)	1.5	(3.1)	.4	(2.1)	.23*	.06	.31*	-.03	.06	.10	.39*
McDonalds	Burger King	240	.24	(.41)	.2	(2.4)	.2	(1.6)	.22*	.01	.51*	-.09	.07	.07	.03
Whites	Asians	255	.26	(.42)	.4	(1.9)	.4	(2.0)	.21*	.13	.34*	-.03	.06	.07	.11
Flexible	Stable	206	.05	(.48)	.4	(2.0)	-1.2	(2.1)	.20	.05	.35*	-.18	.09	.15	.48*
Emotions	Reason	178	.25	(.43)	-.2	(2.2)	-.2	(2.1)	.19	.12	.36*	-.09	.09	.11	.22
Freedom	Security	195	.31	(.39)	1.2	(2.0)	.6	(2.0)	.18	.04	.09	.03	.09	.11	.26
Family	Career	197	.49	(.37)	1.0	(2.2)	.4	(2.3)	.16	.10	.24*	-.19	.06	.12	.40*
Young People	Old People	196	.39	(.36)	.2	(1.9)	-.4	(2.3)	.16	.02	.17	-.06	.05	.10	.37*
Short People	Tall People	213	.02	(.49)	-.2	(1.6)	-1.5	(1.6)	.15	.09	.34*	.11	.06	.11	.37*
Rich People	Poor People	240	.68	(.43)	-.7	(2.3)	1.1	(2.8)	.14	.03	.33*	-.11	.13	.21	.37*
Cold	Hot	199	.07	(.58)	-1.5	(3.1)	-1.5	(2.5)	.08	.11	.13	.18	.06	.09	.31*
Females	Males	231	.41	(.44)	.3	(2.0)	-.1	(2.0)	.03	-.10	.12	-.03	.05	.10	.33*
Letters	Numbers	218	.29	(.43)	.9	(2.3)	.2	(1.6)	.02	.13	.13	.08	.07	.10	.20
Skirts	Pants	218	.26	(.42)	-.7	(2.2)	-.8	(1.9)	.00	-.18	.37*	-.25*	.06	.09	.30*
Future	Past	234	.56	(.37)	1.1	(2.6)	.2	(2.0)	-.05	.09	.29*	-.18	.07	.10	.23*

Note: EA=Explicit Attitudes; EK=Evaluative Knowledge; IAT=Implicit Association Test. * $p < .05$.