Available online at www.sciencedirect.com
ScienceDirect
Journal of Experimental Social Psychology 43 (2007) 399–409
Journal of Experimental Social Psychology
www.elsevier.com/locate/jesp

Plausible assumptions, questionable assumptions and post hoc rationalizations: Will the real IAT, please stand up?

Hart Blanton [a,*], James Jaccard [b], Charlene Christie [c], Patricia M. Gonzales [d]

[a] Texas A&M University, Department of Psychology, 4235 TAMU, College Station, TX 77843-4235, USA
[b] Florida International University, Department of Psychology, University Park, Miami, Florida 33199, USA
[c] Indiana University—Purdue University, Department of Psychology, 4601 Central Avenue, Columbus, Indiana 47203, USA
[d] New York State Energy Research and Development Authority, 17 Columbia Circle, Albany, NY 12203-6399, USA

Received 10 October 2006
Available online 27 December 2006

Abstract

In a recent article, we described psychometric limitations to the Implicit Association Test (IAT). These limitations restrict the utility of this measure and render it problematic for testing many psychological theories that posit a causal role for implicit attitudes. Past failures to recognize this may have promoted faulty conclusions in the literature. In a critique of our article, Nosek and Sriram rejected our entire analysis. They asserted that our original article was based on faulty assumptions and argued that the IAT performs nicely when these assumptions are replaced by other, more plausible assumptions. We show that these plausible assumptions have all the hallmarks of post hoc rationalizations. They make little theoretical sense, are buttressed by deceptive statistical practices, contradict statements these same researchers have made in the past and do little to advance research and theory on implicit attitudes. We close by considering the vigor with which IAT researchers have dismissed meaningful criticisms of their measure.

Keywords: Implicit association test; Psychometrics; Theory testing; Validity; Prejudice; Discrimination

In a recent article (Blanton, Jaccard, Gonzales, & Christie, 2006), we pointed out flaws in the Implicit Association Test (IAT), noting that the measure rests on a dubious psychometric foundation and that its design prevents it from testing many psychological theories that are of interest to researchers. In their critique of our article, Nosek and Sriram (2007) stated that our analysis was flawed. They suggested that we invoked faulty assumptions and asserted that the IAT becomes psychometrically and theoretically sound if one replaces our faulty assumptions with assumptions that they deem to be plausible. We show here, however, that these plausible assumptions have properties suggesting they are post hoc rationalizations: They make little theoretical sense and contradict statements these same researchers have made in other contexts. We also show how Nosek and Sriram seem to lose sight of the purpose of measurement by elevating the IAT above its potential utility. Instead of discussing ways that the IAT might be used as a tool to answer important questions, these researchers suggested ways that research programs might be shaped so that the IAT can be used.

Theory testing

Consider some psychological theories that might be of interest to a psychologist. Fig. 1 presents examples using the traditional schematic for representing causal models. The top panel presents the causal model that would be tested by a researcher who wants to predict college students' identification with math from their attitudes towards math and arts. Based on the path coefficients that this researcher

* Corresponding author.
E-mail address: hblanton@psych.tamu.edu (H. Blanton).

0022-1031/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2006.10.019

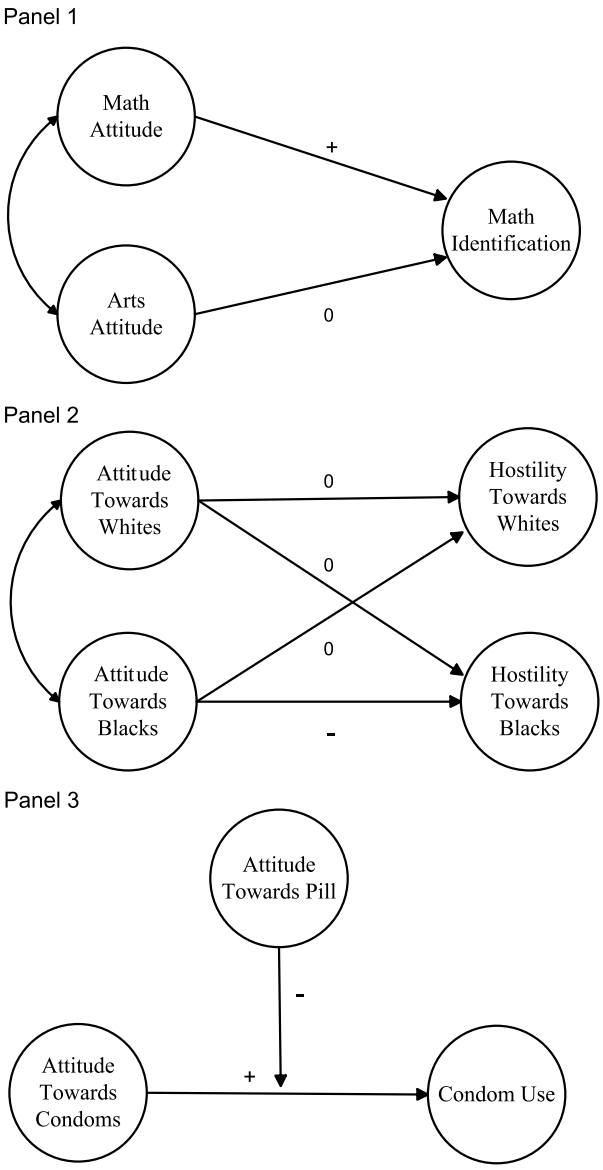

Fig. 1. Causal models using component parts. Note 0 indicates path is set to zero. + Indicates path is a direct relationship. − Indicates path is an inverse relationship.

has hypothesized, math identification should be predicted from the attitude towards math, not the attitude towards arts. This pattern would be consistent with research on attitude-behavior correspondence (Jaccard, 1975; Jaccard, King, & Pomazal, 1977).[1]

The second panel presents a causal model of interest to a researcher who wishes to predict the tendency for white participants to act in a hostile manner towards a white versus a black confederate. The hypothesized path coefficients suggest that this researcher believes that hostility towards a black confederate will be driven by attitudes towards blacks, not attitudes towards whites and that hostility towards whites will be unaffected by attitudes towards blacks. These predictions are consistent with research on attitude-behavior correspondence. It is further hypothesized, however, that correspondence will not be supported in the causal relationship between attitude towards whites and hostility towards whites. Based on research suggesting that members of majority groups do not have strong or consequential attitudes towards their in-groups (Blanton & Christie, 2003; Mullen, 1991), the researcher predicts that attitudes towards whites will not exert significant influence on whites' treatment of a white confederate.

The third panel presents a causal model of interest to a researcher trying to predict adolescent condom use from their attitudes towards condoms and their attitudes towards other methods of pregnancy protection, in this case, the contraceptive pill. The hypothesized path coefficients indicate that a more positive attitude towards condoms will predict greater use of condoms. However, based on past research suggesting that attitudes towards other protection methods are important to consider, the researcher predicts that attitudes toward the pill will moderate the relationship between condom attitudes and condom use: The more positive the attitude toward the pill, the weaker the impact of condom attitudes on condom use behavior.

Each of these theories predicts a reasonably complex causal dynamic based in sound psychological theory. Let's now compare these models with the models that researchers would have to test if they oriented their theories around implicit attitudes and used the IAT as their research tool. These models appear in the three panels in Fig. 2. To avoid the accusation that we have built implausible assumptions into our analyses, we were careful to use the causal form that Nosek and Sriram promoted in their article. Nosek and Sriram proposed that, when testing the causal influence of implicit math and arts attitudes on math identification, a researcher using the IAT would test for the influence of a single IAT factor on this outcome. They labeled this single factor the "latent math-arts attitude" (p. 397). This new attitude construct would replace the two math and arts attitude constructs that were originally of interest to the researcher (see Nosek and Sriram's Fig. 2 for comparison with the first panel of our Fig. 2).

What is a "math-arts attitude" and where did it come from? And, more to the point, why is psychological theory being wrapped around this new construct? On the face of it, the "math-arts attitude" is a construct of questionable value. Metaphorically speaking, it mixes apples with oranges. In the original causal models that we presented in Fig. 1, the causal influence of math attitudes (apples) was assumed to be distinct from the causal influence of arts attitudes (oranges). This makes both common and theoretic sense. But Nosek and Sriram would have psychologists ignore these facts and mix these two psychological constructs together into one big theoretical fruit bowl. This also would occur in the other causal models. In trying to predict

[1] This data pattern would also be consistent with the results that Blanton et al. (2006) reported in their replication and extension of (Nosek, Banaji, & Greenwald, 2002).

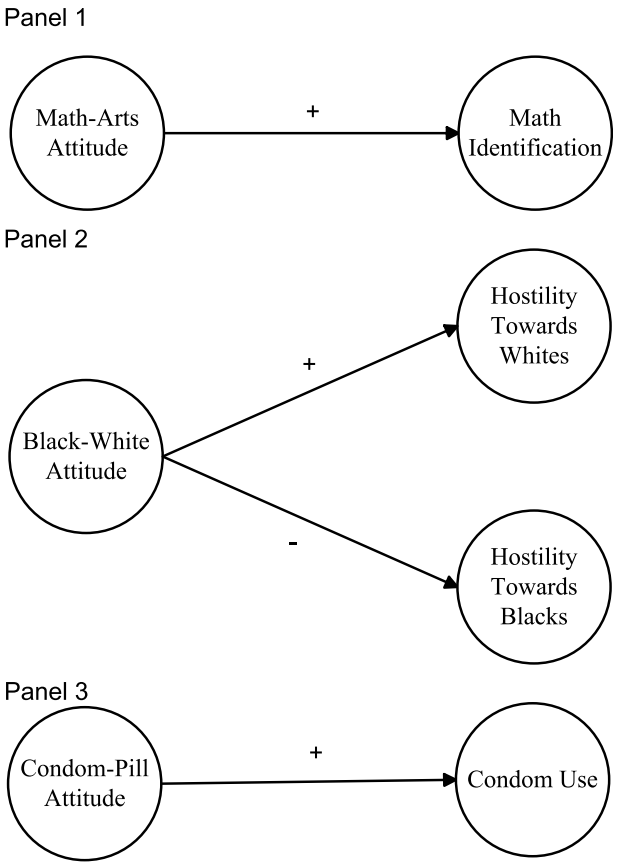

Fig. 2. Causal model using preferences.

the nonverbal treatment of white and black confederates, for instance, a researcher who uses the black-white IAT would not be able to examine the independent influences of black attitudes versus white attitudes. Instead, a researcher would have to adopt a psychological theory that focused attention on a single "black-white attitude." The resulting causal model is represented in the second panel of Fig. 2. Similarly, instead of testing a causal model in which attitudes towards condoms interacts with attitudes towards the pill in the prediction of condom use, a researcher who uses the condom-pill IAT must orient theory around a single "condom-pill attitude." The model that results is shown in the third panel of Fig. 2.

By comparing the models in Fig. 1 with the models that the IAT can test in Fig. 2, it is evident that the IAT is a limited research tool. One goal of our original article was to make this evident to researchers so that they would not apply the IAT in situations where it lacks utility. Instead of exploring the ramifications of the limited causal model built into the IAT, Nosek and Sriram embraced it.

The irrelevance of a one factor result

Whether one is most interested in testing the causal models in Fig. 1 or the causal models in Fig. 2, the theory one ultimately chooses will have consequences for the type of attitude measures one must use. Consider the causal model in the first panel of Fig. 1. This model requires the measurement of two distinct attitudes or, stated another way, it requires attitude measures that have a two-factor structure. One set of items should provide a valid estimate of each participant's "math attitude" and the other set of items should provide a valid estimate of each participant's "arts attitude." If this two-factor structure is not observed for the attitude measures, then they cannot be used to test the model of interest. Now consider the causal model in the top panel of Fig. 2. This model requires an attitude measure that has a single-factor structure. If a single-factor structure is not observed, then the measure is reflecting more than the single construct of interest and this is potentially problematic.

In Blanton et al. (2006), we argued that IAT researchers have ignored the psychometric structure of the component parts of the IAT and instead have moved directly to testing the relatively crude causal model that was built into it. With this in mind, we evaluated the psychometric properties of the component parts of the IAT to determine if the factor structure was consistent with the psychometric model we thought should underlie it. We found that the two components had a multidimensional factor structure, which was inconsistent with the Nosek and Sriram causal model.

Nosek and Sriram objected our analysis and argued that the psychometric model we invoked was not appropriate for the IAT. Ostensibly, this was because the act of computing a difference score from the IAT components creates a meaningful and somewhat mystical aspect of relative attitudes (or preference) that cannot be captured from the component parts of the IAT. They re-scored our data using their *D* method (as we discuss later), then formed multiple instantiations of the difference scores, and showed that these multiple instantiations mapped onto a one-factor model. In essence, they turned a multi-factor solution into a single-factor solution. In the process, they also argued that something special was being captured by the differenced IAT indicators. Although we are heartened that our article has forced IAT theorists to be somewhat more explicit about their psychometric assumptions, we show below that their algebraic manipulations reveal nothing special and that all they have succeeded in doing is mixing apples and oranges.

Nosek and Sriram's justification: The whole is greater than the sum of its parts

Nosek and Sriram's justification for treating their data in the way they did revolved around a hypothesized property of a theoretical construct they called a *relative preference*. They argued that, in many instances, a relative preference cannot be understood by studying the component attitudes that make up this preference. As they said, "relative preferences do not necessarily reduce to component attitudes." (p. 393) In other words, the relative-preference whole is greater than the sum of its attitu-

dinal parts. With this justification in hand, Nosek and Sriram performed the data manipulations that mixed apples with oranges and found what they were looking for—that math-arts IAT has a unidimensional factor structure. As noted, they named this factor a "latent math-arts attitude."

One of the more troubling features of this logic is that it is in direct contradiction to the way IAT researchers have treated relative preferences in the past. This suggests that this response is but a post hoc justification. To begin, suppose that global relative preference measures do have some special quality about them. As a result, there will be times in which relative preferences cannot be detected by measurement strategies that focus on the component parts that make up these relative preferences. Assuming this is true—and assuming that awareness of this fact is one reason that the IAT originally was designed to measure relative preferences—then IAT researchers certainly would have worked under this philosophy. But they have not. In fact, IAT researchers routinely estimate relative attitudes by focusing their analyses on the component parts that make up a relative preference. And there is no greater offender than Nosek.

We examined every empirical article that Nosek has published (as retrieved by PsychInfo). In *every* article that had a measure of explicit attitudes, he pursued a measurement strategy (in one or more instances) whereby the component attitudes underlying explicit relative preferences were measured *separately* and then differenced to create a single estimate of explicit relative preference. Moreover, these computed difference scores were presented by Nosek as the conceptual equivalent of the relative preference assessed by the IAT. As one example, Nosek (2005) described his strategy for assessing relative explicit attitudes as follows:

> "Explicit preferences were assessed by calculating the difference between feelings of warmth ratings for the two social objects *to conceptually parallel the relative measurement feature inherent in the IAT*." (p. 570; italics added)

This statement could not be more clear. In fact, it mirrors statements in Nosek et al. (2002), the article that we critiqued in our original article. Recall that our analyses focused on the math-arts IAT. And recall that Nosek and Sriram argued that this attitude could not be understood if it is reduced to its component parts. Now consider Nosek, Greenwald and Banaji's original statements about their strategy for measuring the explicit math-arts preference:

> "To assess explicit attitudes toward math and arts, we had participants complete paper-and-pencil questionnaires. Specifically, we used feeling thermometers (preference ratings based on a 0–100 scale from cold/unfavorable to warm/favorable) to assess participants' feelings of warmth toward math and arts as academic domains. *By taking the difference between the math and arts temperature ratings, we made the explicit attitude measures comparable to the implicit measures.*" (p. 48; italics added)

In both of these articles, Nosek conceptualized an explicit relative preference as something that could be understood by examination of it component parts.[2] But then Nosek and Sriram took us to task for doing exactly the same thing. In fact, this objection was the primary reason Nosek and Sriram rejected the results of our second study. This study measured the component parts of implicit attitudes and showed a case in which implicit attitude structure does not support the IAT measurement strategy.

If relative attitudinal preferences have some mysterious aspect to them that cannot be assessed by measuring the component parts, and if Nosek knew this to be true, then his own work on explicit measures is flawed. By his own logic, psychologists must now call into question his past conclusions regarding the properties of explicit attitudinal preferences. They must also call into question past conclusions he has made regarding the IAT. This is because the component explicit attitude measures were used by Nosek and colleagues as a criterion when they validated a new scoring algorithm for the IAT (Greenwald, Nosek, & Banaji, 2003). But if the criterion measures were invalid, then this casts doubt on the new scoring algorithm.

Intransitive preferences

It is not uncommon for IAT advocates to make sweeping statements about precedent in other areas of psychology and then to suggest that IAT critics lack knowledge or appreciation of these findings (see Banaji, Nosek, & Greenwald, 2004; Greenwald, Nosek, Banaji, & Klauer, 2005; Greenwald, Nosek, & Sriram, 2006; Greenwald, Rudman, Nosek, & Zayas, 2006). There were multiple instances of this in Nosek and Sriram. To justify their view of relative preferences, for instance, they cited phenomena in other domains that seem consistent with the assertion that a general preference often is not predictable from its component attitudes. Space constraints limit our discussion of this issue, but consider the literature they cite on intransitive preferences. This literature is generally concerned with how changes in the contexts in which people rate their preferences can alter their preferences (sometimes dramatically so). Finding such contextual effects says nothing about whether a global preference is a function of its component attitudes. In fact, not one of the studies Nosek and Sriram cite focused on traditional attitudinal constructs, nor did a single study test if a relative preference rating could be predicted from its component parts.

[2] The IAT demonstration webpages also assess the component attitudes separately.

An empirical evaluation

We now present data that explores issues central to Nosek and Sriram's critique. In a study using a convenience sample of 132 male and female college students, we administered an attitude inventory that measured two attitudes that a consumer psychologist might target with an IAT task. These were (1) attitudes towards apples and (2) attitudes towards oranges. The attitude towards apples was measured with three (11-point) items and the attitude towards oranges was measured with three separate (11-point) items (see the Appendix for all measures used). An attitude towards apples was operationalized as the sum of the three apple items (alpha coefficient = 0.91) and an attitude towards oranges was operationalized as the sum of the three orange items (alpha coefficient = 0.91). In addition, we obtained a direct global preference rating for apples relative to oranges, in which participants rated on one scale how much they preferred apples relative to oranges or vice versa. The goal of the study was to predict liking of fruit products (which might be of interest to applied researchers who wish to increase fruit consumption). The psychological criteria we sought to predict from attitudes focused on the liking of three different fruit-related products. These were, (1) liking of orange marmalade, (2) liking of applesauce, and (3) liking of fruit salad.

Predicting preference from component attitudes

According to the whole-is-greater-than-the-sum-of-its-parts logic of Nosek and Sriram, the one-item preference rating for apples relative to oranges should not be predictable from (or at least it should not be highly correlated with) the component attitude towards apples and attitude toward oranges. We tested this by regressing the preference rating for apples relative to oranges onto the two individual attitude measures. We observed a multiple correlation of 0.86 ($p < .05$). By any conventional standard, this represents a high degree of association. The regression coefficient for the apple attitude was 0.46 and the regression coefficient for the orange attitude was −0.49, with both being highly statistically significant. This calls into question Nosek and Sriram (2007) assertion that the relative preference judgment captures something special that cannot be addressed with the component attitudes themselves.[3] We present additional data below that further question this assertion.

The factor structure of the six attitudinal items

To explore the factor structure of the six attitude items (three for apples and three for oranges), we conducted a traditional exploratory factor analysis. The factors were extracted using a maximum likelihood algorithm with an oblimin rotation to allow for correlated factors. Examination of the eigenvalues indicated a two-factor solution accounting for 85% of the input variance-covariance. The two extracted factors were correlated 0.16.[4] Table 1 presents the pattern matrix representing the factor loadings. It can be seen that the items segregate well along the lines of the two latent attitudes. This finding is consistent with the two-factor structure shown in Fig. 1 in Nosek and Sriram. The factor loadings for items measuring the same attitude are large on one factor and negligible on the other.

Table 1
Factor loadings for factor analysis of apple–orange items

Question	Factor 1	Factor 2
Liking the taste of oranges	.859	−.036
Favorable feeling towards oranges	.948	.092
Pleasure in orange eating experience	.824	−.036
Liking the taste of apples	−.005	.894
Favorable feeling towards apples	−.003	.937
Pleasure in apple eating experience	.008	.821

Consistent with the difference score approach in Nosek and Sriram, we next formed three difference scores using the six items so as to reflect the preference for one attitude object over another. Specifically, the preference was operationalized as the difference in the attitude item for one object (apples) minus the attitude item for the other object (oranges). We thus created three difference scores (the taste item for apples minus the taste item for oranges, the favorable item for apples minus the favorable item for oranges, and the eating experience item for apples minus the eating experience item for oranges). We factor analyzed these three difference scores along with the global preference item that directly assessed the preference.

As expected, the two-factor structure that was built into the questionnaire was eliminated and a single factor that accounted for 87% of the input variance-covariance emerged. Table 2 presents the factor loadings for this solution. It can be seen that the factor loadings are large and that the items seem to reflect well a single underlying preference. These analyses show that there is nothing special about obtaining a one-factor solution via the Nosek and Sriram scoring method. Even though common sense dictates that attitudes towards apples and attitudes towards oranges are psychologically and psychometrically distinct from one another, the Nosek and Sriram approach effectively mixes apples and oranges. Note also that the single-item global preference rating loaded highly on the same factor. Contrary to Nosek and Sriram, the one factor solution in no way points to the existence of some unique and unspecified aspect of attitudinal preference.

[3] Interestingly, however, these results also support the procedures in Nosek (2005), where relative preferences were estimated by computing a difference between two component attitudes.

[4] Further probes of the data uncovered some evidence for small amounts of method variance in the pairs of items (i.e., the two taste items, the two favorability items, and the two eating experience items) but this method variance was inconsequential and affects none of the conclusions reached from these data.

Table 2
Factor loadings for factor analysis of apple–orange differenced items

Question	Loadings
Difference in liking	.930
Difference in favorability	.976
Difference in pleasure	.863
Relative preference	.873

Predicting criteria from preference versus component attitudes

We next conducted analyses to predict our three criteria (liking of applesauce, liking of orange marmalade, and liking of fruit salad) from the attitude measures. We first tested the causal model that Nosek and Sriram advocated in their Fig. 2 (and presented in our Fig. 2). We then tested a model that respected the two-factor structure of our questionnaire (consistent with our Fig. 1).

The Nosek–Sriram causal model

We first fit a path model predicting the criteria from the relative preference, operationalized by the difference of the two attitude measures.[5] As shown in Fig. 3, the path coefficient from the relative preference to the liking of orange marmalade was statistically significant (unstandardized coefficient $= -.065$, standardized coefficient $= -.24$, $p < .005$).[6] The parameter for this effect was in a direction that suggested that, as preference moved more towards preferring oranges over apples, people were more apt to like orange marmalade. The path coefficient from the preference variable to liking of applesauce was not statistically significant, nor was the path coefficient linking the preference variable to the liking of fruit salad.

Before turning to the causal model that used the two attitudes separately, it is informative to consider how psychologists might react to the results of the previous analysis. The results in Fig. 3 might lead psychologists to develop complex psychological theories that explain why the tendency to prefer oranges over apples was significantly related to the liking of orange marmalade but not significantly related to the liking of applesauce. Researchers might further develop theories to account for the irrelevance of relative preference factor in the liking of fruit salad. Perhaps people do not think of apples or oranges when they think of fruit salad. Instead their minds wander to more exotic fruits, such as kiwis. In short, the mere fact that the IAT invokes a crude causal model in research investigations in no way prevents researchers from developing nuanced theories that seem to account for their findings.

The two-factor model

For this analysis, we fit a causal model that used the separate attitudes to predict the three criteria, based on the model in Fig. 4.[7] The analysis revealed (1) a statistically significant path coefficient between attitudes towards oranges and the liking of orange marmalade but not the liking of apple sauce, (2) a statistically significant path coefficient between attitudes towards apples and the liking of applesauce but not orange marmalade, and (3) a statistically significant contribution of both attitudes for the liking of fruit salad. These results are consistent with common sense and are readily interpreted. They show that (1) liking for orange-related products is predicted by orange attitudes, not apple attitudes, (2) liking for apple-related products is predicted by apple attitudes, not orange attitudes, and (3) liking for multi-fruit products is predicted by both orange and apple attitudes.

Note that the results from these models not only conform to common sense, they also result in stronger criterion prediction than was found using the IAT causal model. Whereas the IAT model generally accounted for 2–3% of the variance in the criteria, the two component approach accounted for about 15% of the variance.[8] Superior prediction is to be expected with the two component model because the IAT model is misspecified, i.e., it assumes equal but opposite influence of the two attitudes (see our original critique).

Global relative preference and unique explained variance

In a final set of analyses, we examined Nosek and Sriram's suggestion that a global preference measure can explain variance in a criterion that can't be captured by the individual attitudinal components. Specifically, we added the global measure of preference as a predictor in the linear equations predicting each criterion from the component attitudes. In no case did it add statistically significant additional explained variance. Thus, there was no evidence that the global preference measure reflected some unspecified, psychologically meaningful construct over and above the component attitudes that aided the prediction of the external criteria.

Conclusions from the empirical evaluation

In sum, the data suggest that when apples and oranges are mixed together in a manner consistent with IAT methodology, one is left with a measurement model of precisely that, mixed apples and oranges. Nosek and Sriram would have researchers believe that a single-factor solution for differenced items somehow reveals a psychological con-

[5] We allowed for correlated errors among the criteria to reflect the obvious fact that the correlations between the criteria are impacted by factors other than simply the IAT inspired preference variable (e.g., liking of fruit in general).

[6] This is a just-identified model, hence fit indices are irrelevant.

[7] As before, this model is just identified and so overall fit indices are not germane.

[8] The squared multiple correlations yielded similar values as those implied by the standardized residuals when calculated using the method of Bentler and Raykov (2000).

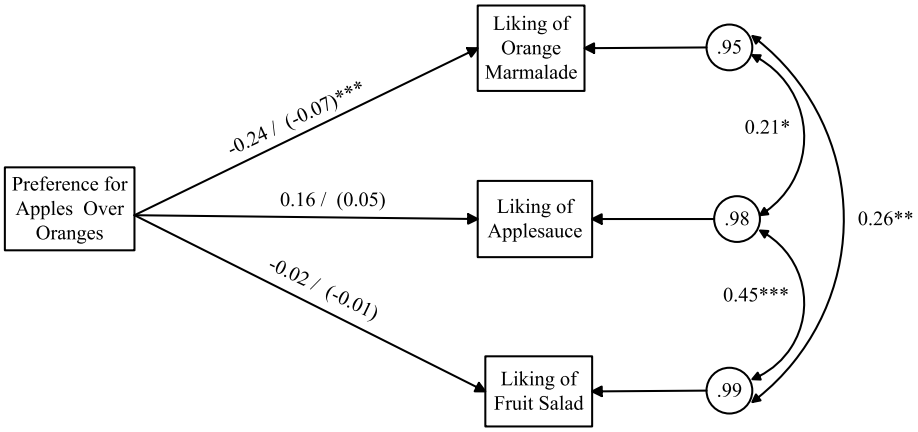

Fig. 3. Path model for the Nosek-Sriram causal structure.

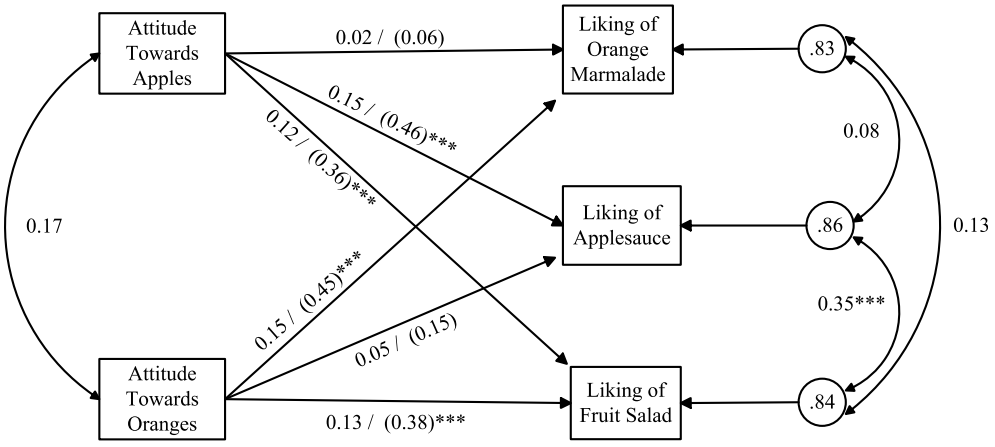

Fig. 4. Path model for the component attitude model.

struct that captures unique aspects of a preference; aspects that cannot be captured by the individual attitudes that comprise the preference. This is not true. We showed that a one-factor solution can be obtained from items that clearly have a two-factor structure by the simple act of differencing. We also showed that the impoverished one-factor model that Nosek and Sriram advanced (1) predicted criteria less well than the two-component model, (2) invoked a relative construct that added nothing meaningful to the prediction of criteria, and (3) yielded a pattern of results that was less theoretically coherent than the analysis of component attitudes. All of these findings support our original critique of the IAT.

Applied consequences of misspecified causal models

Our discussion thus far has focused on the utility of the IAT to test psychological theory, but it also is instructive to consider the applied implications of the Nosek–Sriram conceptualization. Note that in the apples and oranges data, the Nosek and Sriram model yielded a statistically significant path coefficient from the preference for apples over oranges to the liking of orange marmalade. This suggests that, by reducing people's preferences for apples over oranges, one can make them like orange marmalade more. Thus, one can make people like orange marmalade more by making them like apples less, a rather counter-intuitive prediction. As an even more extreme example, Swanson, Rudman, and Greenwald (2001) used an IAT and a set of explicit measures that assessed the preference for smoking over the act of stealing to determine if these measures would discriminate smokers from non-smokers. The explicit measures did, though the IAT was only marginally significant. If one applies the Nosek and Sriram model, it appears that one way to make people smoke less is to reduce their preference for smoking over stealing. This could be done by promoting a more positive attitude towards stealing, while holding constant the attitude towards smoking. Although people might steal more as a result, they should smoke less. A similar result was found in this study in their analysis of the preference for smoking over sweets. The nature of this result was that, if one makes attitudes towards candy more positive, people will smoke less.

These examples make evident that the Nosek–Sriram causal model is quite demanding in the assumptions it embraces. Under the relative-preference framework they

promote, the two attitude objects both are assumed to influence the criterion of interest. Under the simple differencing strategy they embrace, regression parameters are imposed that assume equal but opposite influences. If either assumption does not hold, the model is misspecified and can lead to faulty inference.[9,10]

Reverse-scored IAT indicators: A justified assumption

In the psychometric model we tested for the IAT in our original article, we noted that the two measured IAT response latencies (one for the compatible task and the other for the incompatible task) should be negatively correlated with one another, once systematic confounds due to general processing speed are removed. Accordingly, as one's true preference for whites relative to blacks increases, (1) response latencies for the compatible task of the black-white IAT (pairing whites with positive and blacks with negative) should be sped up, *and* (2) response latencies for the incompatible task (pairing whites with negative and blacks with positive) should be slowed down—all other things being equal. We provided a detailed analysis of the IAT methodology, scoring conventions and past statements in the literature to justify this prediction. Consider just this one quote that seems to justify our assumptions:

> "Shorter reaction times for stimuli belonging to a target concept category which is paired with positive attributes *and* prolongation of the reaction time when this target concept is paired with negative attributes, is interpreted as an indication of a more positive implicit attitude toward this concept category than toward the other concept category with which it is tested." (Maison, Greenwald, & Bruin, 2001, p. 2, italics added)

If the IAT does have these properties, then the two latencies comprising the IAT can indeed be viewed as (reverse-scored) parallel "items" that serve as alternative indicators of the same latent construct, because as one speeds up, the other should slow down. Our original paper tested this model and did not find support for it.

We interpreted this finding as a threat to the assumed psychometric qualities of the IAT, but Nosek and Sriram contested this. They argued that researchers need not assume that the two latencies reflect parallel items with the kind of oppositional dynamics we suggested, and they invoked change scores as an illustration of a special case where differencing does not invoke such dynamics. They stated that "For example, to assess the effectiveness of a math exercise on learning, a math test could be administered before and after the exercise. The conceptual interpretation sought is change in math performance." They suggested that this example shows that "difference scores are not always decomposable as parallel items that separately measure the intended construct," and went on to suggest that we committed "the fallacy of affirming the consequent" by not considering such an obvious exception to the rule we proposed. (p. 395)

As it turns out, we agree with Nosek and Sriram's characterization of change scores. They are correct that the use of a pre-post difference score to reflect change is a special instance in which a researcher does not embrace the type of psychometric model we imposed on the IAT. Interestingly, however, Nosek and Sriram were simply recycling a point that we ourselves had entered into the conversation. Nosek acted as a reviewer for an earlier version of our paper and, in a submission that he read, we included a section exploring special cases where one need not embrace the IAT assumption about difference scoring. One case we explored was pre-post change scores:

> "A third justification for viewing the two IAT response latencies as distinct factors is what we term a co-determinant rationale … Suppose we measure change in a construct by calculating a change score defined as a posttest score minus a pretest score. The underlying latent variable for the measured difference score is true change over time. However, the operative psychometric model does not assume that both measures are alternative indicators of change."

We went on to discuss why this scenario did not apply to the IAT. When analyzing change scores, the two variables that comprise the difference score measure the *same* aspects of the *same* construct. And, they are scored in the same direction (e.g., higher scores mean a more positive attitude). Differences that are observed between the two scores can thus be attributed to external influences (e.g., the effect of an intervention) that cause participants' true (latent) scores to change (plus the effects of measurement and/or nonsystematic error). This dynamic cannot be imported to the IAT. Even a casual analysis shows that the two IAT tasks do not measure the same construct, scored in the same direction, nor have researchers asserted that scores on the two IAT tasks differ from one another because participants' implicit attitudes have changed between assessments on the two tasks. To the contrary—and as discussed in detail in our paper—the two IAT tasks were designed in such a way that they seem to measure opposing aspects of the same (unchanged) implicit relative attitude.

To us, everything in the extant IAT literature suggested a tendency to assume the oppositional dynamics (as quoted above). Indeed, the compatible task functionally becomes the incompatible task for those who hold a

[9] For a discussion of functions other than differencing relating the component attitudes to the global preference, see Jaccard and Blanton (2006).

[10] In our original paper, the Simple Association Test was based on a psychometric model where the two attitude objects were not assumed to be complementary, but rather one was chosen as a measurement standard that acted as a constant to assist in the isolation of the component associations. Thus, there are no consequent change implications for the attitude object that was chosen to serve as a constant.

reversed preference (e.g., for the person who prefers blacks over whites). Yet Nosek and Sriram rejected the reverse scored-parallel item viewpoint. Such is their prerogative. But it is not sufficient simply to make passing reference to change scores and then to suggest that this justifies IAT scoring conventions. A more detailed psychometric analysis is called for, one that explains how the latent preference variable maps onto the four associations that are measured by the two response latencies of the IAT (see our original Fig. 1) and then converted into a single IAT score. Instead of providing such details, Nosek and Sriram made passing reference to change scores and suggested that we were intellectually flawed for not thinking of this example for ourselves. They pursued this strategy, even though the first author knew that we had given careful thought and consideration to this very analogy, as well as others.

Distractions

At the risk of being drawn into nonproductive exchanges, we feel we should respond to two of the false and misleading statements Nosek and Sriram made concerning our original analyses. Each of these statements represents a distraction that might draw readers' attention away from the more fundamental intellectual differences between us.

Distraction 1: The IAT Is Not Double-Barreled. In our original article, we observed that the IAT question format is double-barreled. Specifically, we stated that

> "For the race IAT, response latencies on just the compatible task are influenced by two associations, (1) the association between Whites and positive concepts and (2) the association between Blacks and negative concepts. This IAT 'item' is thus double-barreled in the sense it is influenced by two distinct association strengths … Similar arguments can be made for the incompatible task." (p. 204)

Nosek and Sriram rejected this argument. They failed to mention to readers, however, that Nosek himself embraced this very characterization of the IAT when he had a measure of his own to promote, the Go/No-Go task. When Nosek and Banaji (2001) introduced their Go/No-Go task, they had this to say about the IAT question format:

> "Because [IAT] responses to items belonging to one pairing (*white* + *good*) occur at the same time as responses to other pairings (*black* + *bad*), faster response times are a product of both associate pairs. That is, what may appear to be an association between white and good could also be a function of a strong association in the opposing pairing (*black* + *bad*) or even inhibition from the categories requiring opposite responses (e.g., *black* and *good*)." (p. 631)

We are in complete agreement with Nosek and Banaji (2001) on this point and emphasize again: the IAT is double-barreled in the sense that each item assesses two potentially very different association strengths with a single response latency. The function relating these association strengths to the response latency is unknown.

Distraction 2: The new scoring algorithm fixes everything. We discussed the new IAT scoring algorithm in some depth in our original article, so we will not belabor its limitations here (see our original Appendix A in Blanton, Jaccard, Gonzales, & Christie, 2006). We do note, however, that Nosek and Sriram's invoked this new scoring algorithm to defend the research findings in Nosek et al. (2002)—even though that article did not use the new scoring algorithm in the analyses. This sort of inconsistency is common. IAT researchers often extol the virtues of the new scoring procedure when their older research is criticized, arguing that if it had been used, the results may have been different (see Greenwald et al., 2006; Greenwald, Rudman et al., 2006). But they are quick to cite older research using a different scoring method when it seems supportive.[11]

Nosek and Sriram followed this tradition, but they pursued argumentation that suggests another post hoc rationalization. They claimed that the *D* score is not a difference score. Specifically, they stated that although "it may be tempting to translate *D* into the difference of the two components … this translation is illusory." (p. 396) This argument defies logic: The *D* score is the algebraic equivalent of subtracting the standardized compatible score from the standardized incompatible score. It is mathematically equivalent to a difference score. End of story. In fact, even IAT researchers introduced this scoring convention as nothing more than an alternative means of computing a difference. In the article that introduced the *D* score, Greenwald et al. (2003) noted that the IAT focused on a score that was "conventionally computed as the difference between central tendency measures." (p. 200) They then examined the properties of the *D* score as part of a larger investigation of "five [new] candidate methods of computing this difference." (p. 200). Although Nosek and Sriram refused to concede even this point, an effect size is fundamentally a difference score. Our article explored the reasons that this *D* score is suboptimal.

Closing thoughts

We have now published three critiques of the IAT (Blanton & Jaccard, 2006a, 2006b; Blanton et al., 2006), but each of these papers has been categorically rejected by IAT researchers (Greenwald et al., 2006; Greenwald, Rudman et al., 2006; Nosek & Sriram, 2007). Although we are satisfied with how we have defended our work (Blanton & Jaccard, 2006c, 2006d; Jaccard & Blanton,

[11] The vast majority of studies use an older scoring method that IAT researchers now reject.

Appendix A

Apples and oranges study

Apple attitude questions

In general, how much do you like the taste of *apples*?

0	1	2	3	4	5	6	7	8	9	10
Not at all			Slightly			Quite a bit			Extremely	

In general, how favorable do you feel toward *apples*?

0	1	2	3	4	5	6	7	8	9	10
Not at all			Slightly			Quite a bit			Extremely	

In general, how pleasant do you find the *apple eating experience*?

0	1	2	3	4	5	6	7	8	9	10
Not at all			Slightly			Quite a bit			Extremely	

Orange attitude questions

In general, how much do you like the taste of *oranges*?

0	1	2	3	4	5	6	7	8	9	10
Not at all			Slightly			Quite a bit			Extremely	

In general, how favorable do you feel toward *oranges*?

0	1	2	3	4	5	6	7	8	9	10
Not at all			Slightly			Quite a bit			Extremely	

In general, how pleasant do you find the *orange eating experience*?

0	1	2	3	4	5	6	7	8	9	10
Not at all			Slightly			Quite a bit			Extremely	

Apple–orange preference question

To what extent do you prefer apples over oranges?

−10	−9	−8	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8	9	10
Extreme preference for oranges				Quite a preference for oranges			Slight preference for oranges			No preference			Slight preference for apples			Quite a preference for apples			Extreme preference for apples	

Criteria

How much do like *orange marmalade* (a jelly made with oranges)?

0	1	2	3	4	5	6	7	8	9	10
Not at all			Slightly			Quite a bit			Extremely	

How much do like *apple sauce*?

0	1	2	3	4	5	6	7	8	9	10
Not at all			Slightly			Quite a bit			Extremely	

How much do like *fruit salad*?

0	1	2	3	4	5	6	7	8	9	10
Not at all			Slightly			Quite a bit			Extremely	

2006), we cannot help but notice that we have had little influence on how IAT research currently is practiced. Instead of finding creative ways of improving the IAT or addressing our concerns, the developers of the IAT instead have sought creative ways of defending their original positions. Creativity of this sort is not unfamiliar to science. Numerous methodologists have pointed out that a clever theorist can invent auxiliary assumptions that circumvent most any criticism. Indeed, the issue of adherence to dubious phenomena was central in an analysis by one of the leading IAT theorists, when he critiqued social psychology's tendency to embrace the sleeper effect (Greenwald, Pratkanis, Leippe, & Baumgardner, 1986). Twenty years later, IAT advocates now maintain a "validity debates" webpage that promotes the auxiliary assumptions needed to keep faith in the IAT.[12] This webpage reviews most of the strong criticism of the IAT (including our own), but then dismisses each of them as minor points that are of little importance.

With this in mind, we suspect that IAT advocates will respond to the empirical data reported here by suggesting that, although there are some cases in which a preference is predictable from its component parts, there also are cases in which it is not—and we just happened to stumble upon an instance of the former. Or, perhaps they will suggest that the empirical data we reported were operating at the level of explicit phenomena, but that things are somehow different at the implicit level. Despite the ongoing process of adding auxiliary assumptions to protect the IAT, we believe that science is a self-correcting process. Ultimately, the value of the IAT will be made evident.

References

Bentler, P. M., & Raykov, T. (2000). On measures of explained variance in nonrecursive structural equation models. *Journal of Applied Psychology, 85*(3), 330.

Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2004). No place for nostalgia in science. A response to Arkes and Tetlock. *Psychological Inquiry, 15*, 279–289.

Blanton, H., & Christie, C. (2003). Deviance regulation: a theory of identity and action. *Review of General Psychology, 7*(2), 115–149.

Blanton, H., & Jaccard, J. (2006a). Tests of multiplicative models in psychology: a case study using the unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 113*, 155–169.

Blanton, H., & Jaccard, J. (2006b). Arbitrary metrics in psychology. *American Psychologist, 61*, 27–41.

Blanton, H., & Jaccard, J. (2006c). Postscript: perspectives on the reply by Greenwald, Rudman, Nosek and Zayas (2006). *Psychological Review, 113*(1), 166–169.

Blanton, H., & Jaccard, J. (2006d). Arbitrary metrics redux. *American Psychologist, 61*(1), 62–71.

Blanton, H., Jaccard, J., Gonzales, P., & Christie, C. (2006). Decoding the implicit association test: Perspectives on criterion prediction. *Journal of Experimental Social Psychology, 42*(2), 192–212.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197–216.

Greenwald, A. G., Nosek, B. A., Banaji, M. R., & Klauer, K. C. (2005). Validity of the salience asymmetry interpretation of the IAT: Comment on Rothermund and Wentura (2004). *Journal of Experimental Psychology: General, 134*, 420–425.

Greenwald, A. G., Nosek, B. A., & Sriram, N. (2006). Consequential validity of the Implicit Association Test: Comment on the article by Blanton and Jaccard. *American Psychologist, 61*, 56–61.

Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review, 93*, 216–229.

Greenwald, A. G., Rudman, L. A., Nosek, B. A., & Zayas, V. (2006). Why so little faith? A reply to Blanton and Jaccard's (2006) skeptical view of testing pure multiplicative theories. *Psychological Review, 113*, 170–180.

Jaccard, J. (1975). A theoretical analysis of selected factors important to health education strategies. *Health Education Monographs, 3*, 152–166.

Jaccard, J., & Blanton, H. (2006; in press). A theory of implicit reasoned action: the role of implicit and explicit attitudes in the prediction of behavior. To appear in I. Ajzen, D. Albarracin, & J. Hornik. *Prediction and change of health behavior: Applying the reasoned action approach*. Mahwah, New Jersey: Erlbaum.

Jaccard, J., King, G. W., & Pomazal, R. P. (1977). Attitudes and behavior: an analysis of specificity of attitudinal predictors. *Human Relations, 30*, 817–824.

Maison, D., Greenwald, A. G., & Bruin, R. (2001). The Implicit Association Test as a measure of implicit consumer attitudes. *Polish Psychological Bulletin, 2*, 61–79.

Mullen, B. (1991). Group composition, salience and cognitive representation: the phenomenology of being in a group. *Journal of Experimental Social Psychology, 27*, 297–323.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*(4), 565–584.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*(6), 625–666.

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

Nosek, B. A., & Sriram, N. (2007). Faulty assumptions: A comment on Blanton, Jaccard, Gonzales and Christie (2006). *Journal of Experimental Social Psychology, 43*, 393–398.

Swanson, J. E., Rudman, L. A., & Greenwald, A. G. (2001). Using the Implicit Association Test to investigate attitude-behavior consistency for stigmatized behavior. *Cognition and Emotion, 15*, 207–230.

[12] http://faculty.washington.edu/agg/iat_validity.htm.