
Summary of Nosek and Sriram (2007) comment

The opening statement from Nosek and Sriram (2007, hereafter “Comment”) was: “This article makes two points: (1) BJGC are too quick to dismiss the value of relative measurement and mistakenly suggest that relative attitude measures are necessarily reducible to component evaluations; and (2) BJGC misperceive the Implicit Association Test (IAT; Nosek, Greenwald, & Banaji, 2006) conditions as analogous to two items of a scale and consequently impose invalid psychometric assumptions by decomposing the performance conditions as separate indicators of the construct.”

Summary of Blanton, Jaccard, Gonzales, and Christie (2007) reply

In their reply (hereafter “Reply”), BJGC address issues that are not relevant to the points of contention, and, most critically, forget the distinction between the conceptual level (whether the attitude objects in the IAT can be represented as an additive difference) and the observed level (model of the IAT effect based on its two response conditions, which do not correspond with the single attitudes). This Postscript summarizes the most critical errors.

Nosek & Sriram’s (2007), Point (1): BJGC are too quick to dismiss the value of relative measurement and mistakenly suggest that relative attitude measures are necessarily reducible to component evaluations

The Comment took issue with BJGC’s statement that “Because the IAT is focused on relative implicit evaluations at the conceptual level, theorists who use the IAT to predict psychological criteria must assume that the implicit evaluations of two distinct objects combine in additive fashion to impact the criterion of interest.” (BJGC’s original article, page 207, hereafter [Article]). The Comment rejected the assertion that relative attitude measures are necessarily reducible to component evaluations, and that this is an inherent property of the IAT. *Whether it is true or not is a theoretical commitment that can be resolved empirically.*

In the Reply, though it is not stated as such, BJGC appear to agree that the Article’s language was too strong because the argument strategy shifts from claiming that this is something that one *must assume,* to making a case that it is true by precedent, by reference to Nosek’s previous publications, and making a hasty generalization from an empirical example. The Reply also misstates the Comment’s argument by suggesting that it rejected the possibility that relative preferences could be decomposed as additive differences. For example, “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” (Reply). The Comment expressed no discomfort with the possibility that the additive difference model *could* be true (in fact, supplementary materials report evidence that this claim holds for the [Article’s Study 1 data]). Nosek and Sriram’s criticism was that there is no justification to claim that it *must* be true.

Confusing conceptual and observed levels

The Reply makes a critical error by blending what BJGC’s Article referred to as the *conceptual level* versus observed level of the IAT. In that paper, BJGC acknowledged correctly that the observed level “is quite different from the one at the conceptual level” (BJGC, page 194). The Reply forgets that the two IAT conditions do not correspond to the component attitude objects, leading to an inaccurate analysis.
The Reply’s lengthy discussion about the conceptual level (i.e., whether relative preferences in the IAT can be characterized as the additive difference of two component attitudes) has no implications for the structural models reported in the Article or Comment. The conceptual level concerns the attitude objects – whether and how they are decomposable into single attitudes. The observed level concerns the calculation of an IAT effect from the two response conditions (e.g., [1] Black+bad, White+good; [2] Black+good, White+bad). Some examples of the conflation of these in the Reply:

- The structural models examined in the Reply used single attitudes as latent factors, whereas the models in the Article and Comment represented the IAT response conditions.

- As motivator for a section about whether relative attitudes can be decomposed (conceptual level) the Reply states “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.”, which is a comment on the observed level – the IAT response conditions.

- The Article and Comment models were about combining IAT conditions (observed level), the Reply models were about component attitudes (conceptual level). The Reply equated them throughout, e.g., “These analyses show that there is nothing special about obtaining a one-factor solution via the Nosek and Sriram scoring method” (ms page 12). The models were not investigating the same question.

- Another instance of this error is also an example of affirming the consequent. The fact that a relative preference does not necessarily reduce to component attitudes (the Comment’s conceptual claim), does not imply the Reply’s inaccurate characterization of the Comment’s claim: “Nosek and Sriram would have researchers believe that observing a single factor structure for differenced items somehow reflects a construct that captures unique aspects of a preference; aspects that cannot be captured by the individual attitudes that comprise the preference.” (Msp 15).

- Again, that the two conditions of the IAT (observed level) do not correspond to the component attitude objects or how they are combined (conceptual level) is missed in the Reply’s statement that “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.” (Msp 16).

The Article and Comment made clear distinctions between conceptual and observed levels. The Reply did not introducing considerable confusion and leading to mischaracterization of the Comment’s claims. When the confusion is removed, all that remains of the Reply are claims that (1) relative measurement has little scientific value, and (2) there are occasions in which relative preferences can be reduced to their component attitudes. We disagree with the first point, and agree with the second (and never stated otherwise). We even agree that the second point could be true of the IAT (see the other supplements). We disagree that it must be true.

Nosek & Sriram (2007), Point (2): BJGC misperceive the IAT conditions as analogous to two items of a scale and consequently impose invalid psychometric assumptions (parallel items versus condition contrasts)

The Comment’s Point (1) was a criticism of the Article’s unjustified conceptual level commitment. Point (2) was a criticism of the Article’s imposed assumptions at the observed level.
The Reply acknowledges that change scores cannot be treated as separate indicators of a single construct, and so agree that their models do not apply to all applications of difference scores (msp 22). The Reply states:

“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.”

This is useful because it acknowledges that the construct assessed by a difference score at the observed level is not necessarily decomposable into component parts, something that the Article stated must be true. However, the Reply rejects that the change score logic can be applied to the IAT saying:

“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.”

This statement is correct about the conditions necessary for establishing a change score, and incorrect in claiming that this does not apply to the IAT. The Comment pointed out that this logic applies to the IAT and most other contrasted performance paradigms. As reflected in the Reply quote above, in many cases, the construct of interest is the change in value on a DV in response to some sort of manipulation. The IAT and other tasks that we reviewed are exactly this – manipulations of response conditions to measure their effect on a dependent variable. Here are a few examples from the Comment:

<table>
<thead>
<tr>
<th>Content of interest</th>
<th>Manipulation</th>
<th>DV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math learning</td>
<td>Educational Intervention (pre or post)</td>
<td>Number of correct responses on a math test</td>
</tr>
<tr>
<td>Stroop interference</td>
<td>Word-color (match or mismatch)</td>
<td>Latency of color identification</td>
</tr>
<tr>
<td>Evaluative priming effect</td>
<td>Prime (positive or negative)</td>
<td>Latency of evaluating a target</td>
</tr>
<tr>
<td>IAT effect</td>
<td>Response key configuration (‘Black with good/White with bad’ or ‘Black with bad/White with good’)</td>
<td>Latency of categorizing a target</td>
</tr>
</tbody>
</table>

In all these cases, the change score provides an estimate of the magnitude of the influence of the manipulation on the DV. And, in all cases, the two conditions are “scored in the same direction.” In the IAT, for example, the construct is strength of association. In both conditions, faster responses indicate stronger associations. The difference between conditions is an indicator of which concept combinations are more strongly associated.

The last phrase of the Reply’s quote above “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” is illustrative of how the Reply misunderstands the application of change score logic to contrasted performance paradigms. The attitude has not changed, the response configuration has changed. The IAT components are not direct indicators of attitude, the attitude construct is derived from the comparison of conditions; likewise the two math tests are not direct indicators of learning, learning is derived from the comparison of conditions.

It appears that the Reply would agree with the Comment’s rejection of their models if the interpretation of change scores (contrasted conditions) applies to the IAT. Therefore, the erroneous parallel items analogy – the main point of the Comment – is the core point of contention. Ironically, the original Article concluded that “our [final] model tests suggested an alternative measurement model that treats the ‘compatible’ and ‘incompatible’ IAT judgments as distinct psychological constructs that can have different influences on psychological criteria” (p. 204). This conclusion directly rejects their assumption that the IAT conditions are parallel-items measuring the same latent construct. In summary, BJGC falsify their own assumption and simultaneously refuse to relinquish it because it is the core motivating factor of their critique.

Summary

The Reply mounts a spirited defense of a set of assumptions that only BJGC are willing to endorse, and that they then show to be incorrect. In the Comment we agree that the assumptions are incorrect, explain why, and show that good fitting models are obtained when appropriate assumptions are used. Instead of building on this insight, the Reply redefines the assumptions that they introduced and rejected in the original Article.

Other points

* BJGC dismiss the D scoring algorithm without recognizing that it is not decomposable as a difference between the IAT conditions. The Reply rejects this claim saying “This argument [from the Comment] defies logic: The D score is the algebraic equivalent of subtracting the standardized compatible score from the standardized incompatible score.” The two “scores” do not correspond to the compatible and incompatible conditions because the standard deviation contains data from both conditions – i.e., the two “scores” are interdependent and cannot be represented in structural models without violating the assumption of independent indicators. The Comment elaborates this point.

* The Reply states: “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.” (Reply, ms page 2). What the Comment actually said was “Selection of assessment – whether single or relative – should conform to the goals of the research question.” (msp. 5).

* The Reply says that “this objection was the primary reason Nosek and Sriram rejected the results of our second study.” The Comment never mentioned Study 2 or said anything that would reject it.