Dear Prof Hippel,

First and foremost, I would like to thank both you and Reviewers 1-2 for your highly constructive feedback. Your comments proved highly informative and have helped us to craft an even stronger version of our original manuscript.

Although you opted to reject our original manuscript (based largely on Reviewer 1’s comments) we would like to ask that you reconsider our revised manuscript for several important reasons. First and foremost, we believe that many of Reviewer 1’s concerns stemmed from our attempt to fit our manuscript into the Short Report format rather than fundamental issues with the paper itself. Upon revision, what became clear was that greater space was required to do our findings justice. To this end, we would like to re-submit our manuscript in full Research Article format that has been improved based on the feedback provided.

This new manuscript is superior to our original in several respects, and contains many innovative aspects. For instance, we now include (a) additional meta analyses that synthesise the results from the explicit measures, (b) entirely new analyses that employ innovative statistical techniques to address the conceptual concerns of Reviewer 1, and (c) an in-depth treatment of the potential mechanisms underpinning our effects, the relationship between our findings and key papers in the literature (Vorauer, 2012), as well as a detailed account of unexpected results where they arise.

In particular, the new statistical techniques we used (the assessment of measurement invariance) has widely recognised utility, but has never been employed within the implicit measurement literature. We demonstrate that it provides a solution to the thorny issue of whether changes in scores on an implicit measure represent ‘genuine’ changes in implicit attitudes or not. Beyond its utility to our current research goals, our manuscript describes how others can make use of this method to address this common question in their own work.

We believe that these revisions adequately address the concerns of Reviewers 1-2 and further strengthen an already innovative paper showcasing the various roles that measurement reactivity plays in implicit and explicit measures.

We would like to ask that our paper be given consideration once more - either as a revised manuscript or as an entirely new submission.

Please find below a point-by-point list of how we have addressed changes in the manuscript. As well as these specific points, the introduction and discussion sections have been extensively rewritten to flesh out our arguments.

Once again we thank you for your feedback and input on this manuscript and look forward to your reply.

Best wishes,

Ian and Jan

**Editor’s decision letter**

10-Jan-2019  
  
Dear Dr. Hussey:  
  
Thank you for submitting "Completing a Race IAT increases implicit racial bias" to Psychological Science. I was fortunate to get two thoughtful reviews from experts in attitudes and implicit social cognition. As you can see, both reviewers think you are tackling a very interesting problem that is likely to attract a great deal of attention, and I agree. At that point, however, their evaluations diverge. Reviewer 2 suggests that with some relatively straightforward revisions the paper would be suitable for Psychological Science, but Reviewer 1 argues that although the work should definitely be published, it is inappropriate for this outlet. Reviewer 1's primary concern is that we have no idea why your effect emerged. Given the large number of possible mechanisms, and the lack of data supporting your preferred interpretation, I'm afraid that I found Reviewer 1's concerns compelling. Thus, despite the fact that I think the paper should be published, and despite the fact that I think it will prove to be generative and highly cited, I don't think we are the appropriate outlet for the work. For this reason, I have decided to decline your manuscript for publication in Psychological Science.  
  
I regret that the outcome was not positive, but we have to be extraordinarily selective due to the large number of manuscripts that are submitted to the journal (nearly 2,000 new submissions are expected this year). I wish you well with this project and I hope the reviews are useful as you revise the manuscript for a specialty journal. Finally, I encourage you to continue to consider Psychological Science as an outlet for your work.  
  
Yours sincerely,  
  
Bill von Hippel  
Associate Editor, Psychological Science  
[psci@psychologicalscience.org](mailto:psci@psychologicalscience.org)

REVIEWER(S)' COMMENTS:  
  
**Reviewer: 1**  
  
**Reviewer 1**: 1. I concur wholeheartedly with the point that the authors make about the potential reactivity associated with completing a race IAT.  I think the field in general is less sensitive to the reactivity of measurement procedures than it should be.  That said, I doubt that the present research can be considered sufficiently definitive to warrant publication in a journal of the stature of Psychological Science.Indeed, I think the work raises considerably more questions than it answers.    
  
First of all, no account is offered as to why the Shooter Bias measure was unaffected by the completion of a race IAT, whereas the other two measures were affected.

**Authors**: We thank Reviewer 2 for his very useful feedback. To clarify our goals and the motivation for each experiment in the manuscript: our original purpose in this paper was to examine if completing a widely used implicit evaluative measure (the IAT) not only captures but alters the very evaluations it seeks to test. After observing this outcome in two studies with different implicit evaluative measures (SC-IAT and AMP), we then sought to push our idea even further. Specifically, we wanted to see if completing an implicit evaluative measure would alter performance on a related but non-evaluative task (Shooter Bias Task) which focuses more on violence than evaluative biases. The fact that the IAT did not alter performance on this latter task is important for two reasons.

First, it places clear boundary conditions on the ‘assessment-as-intervention’ effect we set out to test and informs us about the extent to which that ‘assessment-as-intervention’ effect does or does not generalise across different types of bias. Specifically, it suggests that while completing a Race IAT *does* alter those evaluations one aims to measure it does not bias other forms of automatic behavior such as violence biases. This is both highly useful and diagnostic information.

Second, and to the best of our knowledge, there is currently no literature on the relationship between implicit evaluative and violence biases (i.e., studies that directly compare performance on a race IAT to that of the Shooter Bias Task). We did not presuppose that any such relationship exists, and instead, simply set out to test this for the first time.

Therefore, unlike Reviewer 1, we do not view the absence of a relationship between the IAT and Shooter Bias Task as a reason for concern but rather as an opportunity to examine if any such relationship actually exists in the first place. Results suggest that it does not.

However, with Reviewer 1’s comments in mind, we carried out an additional set of analyses that examined the differential correlations between Race or Control IAT effects and performance on the various outcome tasks. We examined whether:

* race iAT should be associated with the sciat, and more so than the control IAT
* race iAT should be associated with the AMP, and more so than the control IAT
* race iAT is perhaps not strongly associated with the shooter bias task, and not more so than the control IAT
  + If so, then it is not so surprising that the intervention does not work.

We now acknowledge and discusss these important points in the general discussion (p.XX).

**Reviewer 1**: Second, and even more importantly, the research offers no insight into the mechanism that is responsible for the observed effects on the SC-IAT and the AMP. Instead, inferences are drawn in very a direct, but in my view logically unnecessary, fashion that attitudes have changed as a result of completing the IAT.  It seems just as plausible, if not more so, that there has been no attitude change, but that the act of completing the IAT sensitizes participants to the race-related nature of the later tasks and leads them to alter the strategies and the mindset by which they approach those measures.

**Authors**: Reviewer 1 asks whether racial attitudes (as measured by an AMP or SC-IAT) ‘genuinely’ changed as a function of completing an earlier race (relative to control) IAT. In statistical terms, this can be formalised in the following way: is performance on the outcome measure reflective of a genuine change in the underlying latent construct for those in the intervention relative to control condition?

One powerful technique, referred to as testing for ‘measurement invariance’, is widely used outside of the implicit attitudes literature and can help us directly answer this question. When applied to the current paper, such tests can tell us whether or not scores on the outcome (AMP and SC-IAT) tasks still function as measures of implicit racial bias for those in the intervention relative to control condition. If so, then the observed differences between those conditions should be ‘measurement invariant’ and thus attributable to ‘genuine’ changes in implicit racial attitudes. If not, and they are shown to be ‘measurement variant’, then it is possible that Reviewer 1 is correct and “there has been no attitude change”.

When we carried out such tests we found that the former was true: scores on the AMP and SC-IAT were measurement invariant between the intervention and control conditions. Put simply, the effects observed in the intervention condition appear to represent ‘genuine’ changes in attitudes. We now highlight and acknowledge this point in the General Discussion (see p.XX).

On a side note, this question of whether changes in scores on implicit measures represent genuine changes in implicit attitudes is a common and pernicious issue in this research area. To our knowledge, no researchers have ever sought to directly test this question in this manner. This stragegy may therefore enable future work involving implicit measures to address this question more easily.

We return to Reviewer 1’s point regarding potential mechanisms in the next point, in which he or she more fully elaborates the issue.

**Reviewer 1**: Third, and relatedly, I was disturbed that the work by Vorauer (Psych Science, 2012) received no consideration other than a cf. citation in the first paragraph of the introduction. That research demonstrated that completion of a race IAT resulted in negative consequences when White individuals later participated in an interracial interaction. Moreover, the experiments highlighted a mechanism responsible for these adverse effects. The IAT enhanced the Whites’ caution and inhibition regarding racial interactions and this cautious approach limited the positive regard that they communicated to the interaction partner. It is precisely this concern with mechanism that is missing from the present research. Did participants who just completed a race IAT approach the subsequent tasks more cautiously?  Just how was their behavior affected?

**Authors**: Reviewer 1’s questions concerning the potential mechanisms underpinning our effects suggests is that our original material in this regard was insufficient. With this in mind, we added further material on our mechanism (analogical learning), and the previous work that supports this (i.e., Hussey & De Houwer, 2018; see p.XX). In line with Reviewer 1’s request, we also discuss Vorauer (2012)’s paper in greater detail, along with its implications for the interpretation of our own results. This new material highlights the differences between our work and that of Vorauer.

Specifically, our paper focuses on changes to implicit evaluative biases whereas Vorauer’s was interested in changes in interracial interactions. Moreover, Vorauer’s implied mechanisms (e.g., caution, inhibition) certainly differ from those that we outline here (analogical learning), however the two are not mutually exclusive. For instance, increased *interpersonal* caution and inhibition may be a downstream consequence of the *intrapersonal* analogical learning that plays out within the IAT itself. We now unpack these various points in introduction (see p.XX).

**Reviewer 1**: Although less central to my major hesitations about the paper, I found it interesting that two of the four experiments found self-reported ratings of the Black individuals to bemore positive after participants completed the race IAT. These findings were essentially dismissed on p. 13 of the manuscript, but I think they may warrant more attention, largely because the direction of the effect runs counter to the increased negativity that was observed on the implicit measures.  In particular, this disparity leads me to wonder whether the relation between this explicit measure and the implicit measure that constituted the major dependent variable was affected by the completion of a race IAT.

**Authors**: Upon reflection it became clear that, in our original manuscript, we did not adequately synthesize the self-report ratings across our four studies (mainly due to the space limitations created by the short report format). In our revised manuscript we now include a meta-analysis of the self-report data (see p.XX). Results of this meta-analysis reveal that participants in the intervention condition (i.e., who completed a race IAT) showed significantly more positive evaluations of Black individuals than their counterparts in the control condition. As Reviewer 1 indicates, this pattern of findings runs contrary to our original expectations.

We believe that ‘assessment-as-intervention’ effects plays a central role here, regardless of the direction of the effect. Participants in the intervention condition initially encountered a context (race IAT) where they had to emit comparative racial evaluative responses to Black and White individuals (i.e., categorize Black and Whites as good or bad), and to do this in a relatively automatic way (i.e., quickly and without much thought or control). Then, in a subsequent phase, they had to emit further racial responses in a relatively automatic way (either on the AMP, SC-IAT, or Shooter Bias Task). In contrast, participants in the control condition never encountered comparative racial evaluations, as the outcome tasks only ever assessed (implicitly or explicitly) attitudes to Black people.

The comparative nature of the IAT may therefore have sensitized participants to the idea of racial comparisons - particularly given that participants were exclusively White and being asked to evaluate Black people. When later given an opportunity to emit evaluations of the racial outgroup in a slow, intentional, and deliberative manner, their ‘assessment-as-intervention’ effects took the form of overcompensation in terms of more positive evaluations in the intervention relative to control condition (where issues to do with racial comparisons were relatively less salient).

In short, the unexpected direction of change in racial evaluations in the intervention condition could have been driven by repeated exposure to a context where race was repeatedly made salient, and as a result, the participant produced overcompensatory responses towards racial outgroup members when given the time and opportunity to do so (in contrast to the various implicit measures). This provides one explanation for why we saw an impact of the race IAT in one direction at the implicit level (i.e., because in that context people are unable to exert control over their responding) but in another direction at explicit level (where people are able to exert control over their responding).

Critically, however, it is important to note that while the direction of the effect on the explicit measures was opposite to what was predicted, the presence of an evaluative learning effect in either direction underscores our core point: that the act of completing a race IAT serves to not only assess racial evaluations but also change those same evaluations. As such, greater consideration should be given to ‘assessment-as-intervention’ effects effects within the implicit measurement literature.

Our manuscript has now been revised to directly speak to these points in the general discussion (see p.XX).

**Reviewer: 2**  
  
Comments to the Author  
Signed: Yoav Bar-Anan  
  
**Reviewer 2**: 1. The manuscript reports four experiments that tested the effect of completing a White people-Black people Implicit Association Test (in comparison to completing a flowers-insects IAT) on the indirectly and directly measured judgment of Black people. The three experiments that measured the evaluation of Black people found that completing the race IAT increased participants’ negative evaluation of Black people measured by a Single-Category IAT and by the Affect Misattribution Procedure. Another experiment that used the shooter-bias task, did not find any effect of the IAT on performance in that task. Two experiments found no effect of the manipulation on self-reported ratings of Black people’s faces (same faces as those used in the IAT), whereas the two other experiments found more positively evaluation of those faces after the race IAT than after the control IAT.   
  
The authors concluded that the IAT can change people’s automatic evaluation of the groups (or at least one group) included in the IAT. They speculated that the mechanism for this effect is that people learn that completing the IAT encourages people to find an analogy between the two pairs of concepts, and that might subsequently influence people’s beliefs. In this case, the authors speculated that because the White participants associated White people with Good, the IAT encouraged them to pair Black people with Bad, leading to the inference that Black people are bad.

The manuscript reports a finding that a very large audience of scientists and other people would find highly interesting, it presents experimental procedures that many researchers would find useful for their own research, and a theoretical account that is likely to attract attention and lead to further investigations. Although their finding is potential explosive, the authors are very cautious in their interpretation of the results and their implications. The authors should also be praised for providing a very convenient access to all their materials online (e.g., I quickly learned about the results from the wide-format processed datasets). In light of all these positive aspects, I have no doubt that the manuscript would have a positive contribution to the scientific community.

I have a few comments the authors might want to consider for a possible revision, but they are all about omissions from the manuscript, which were probably the result of the word limit imposed by the journal. Still, the authors might consider creating a supplement materials document with a summary of the details readers are most likely to seek. Alternatively, the authors might consider reporting in the methods that “As detailed in the supplement materials (LINK), another experiment directly replicated Experiment 2 and found the same results” and omit any other reference to Experiment 4 in the main text, thus freeing precious space, for further clarifications.

**Authors**: We thank Reviewer 2 for his kind words about our work. Resubmitting the manuscript in full Research Report format has given us additional space to work with.

**Reviewer 2**: 2. It seems wise to use LMM for the statistical analysis. However, readers would benefit from seeing the results based on the typical scores computed for the measures used in this research. Specifically, I would have liked to see means and SDs of the SC-IAT D score and AMP difference score for each condition in each study (in the shooter task, perhaps it is common to report SDT scores).

**Authors**: In line with Reviewer 2’s request, and for illustrative purposes, we now report Means, Standard deviations, and Cohen’s d scores between groups using tthe more commonly used *D* score metric in Experiment 1 (see p.XX).

**Reviewer 2**: 3.It would be useful to indicate the reliability of the measures.

**Authors**: In line with Reviewer 2’s request, we now report the internal reliability of all measures in Table XX.

**Reviewer 2**: 4. The manuscript does not explain what the IAT is. 

**Authors**: The introduction section has been expanded and contains greater reference to the IAT (plus the SC-IAT and AMP) as measures of implicit evaluation. The SOM-R also contain specific details of the task, explaining what the IAT is and looks like (see p XX & SOM-R).

**Reviewer 2**: 5.  P. 6: I assume the measures in the procedure are listed in the order the participants completed them, but this is not explicitly indicated.

**Authors**: We now clarify in the manuscript that, in each experiment, measures are “Listed by order of completion, …” (e.g., p. XX)

**Reviewer 2**: 6. P. 7: The AMP is the Affect Misattribution Procedure not the Affective Misattribution Procedure.

**Authors**: We thank Reviewer 2 for highlighting this and have fixed the error (see p.XX).

**Reviewer 2**: 7. P. 9: The factor “SC-IAT block” was probably the SC-IAT pairing condition, rather than block number. This is not clear. Later in the same paragraph, the authors write “SC-IAT effects differed between the two IAT conditions” and probably mean “The IAT condition moderated the effect of the SC-IAT pairing condition”.

**Authors**: We acknowledge that, although reported correctly, the description of our model was not clear. We have therefore clarified the wording to further aid interpretation: “As hypothesized, SC-IAT effects differed between the two IAT conditions (i.e., there was an interaction effect between SC-IAT block and IAT condition in the prediction of reaction times)” (see p.XX)

**Reviewer 2**: 8. P. 10: The authors wrote “racism as fixed-effect covariate” and probably meant “modern racism score”.

**Authors**: We thank Reviewer 2 for highlighting this and have fixed the error (see p.XX).

**Reviewer 2**: 9. P. 4: I think “Yudkin & Bavel, 2018” should be “Yudkin & Van Bavel, 2018”.

**Authors**: We thank Reviewer 2 for highlighting this and have fixed the error (see p.XX).