**Peer Review and Communication History**

**MS Title**: Cognitive Control and the Implicit Association Test: A Replication of Siegel, Dougherty, and Huber (2012)

**Author Names**: Ryan, William H., Evers, Ellen R. K., Moore, Don A.

**Submitted**: Apr 7, 2021

**Editor First Decision—Revise & Resubmit**

June 13, 2021

Dear Dr. Robey,

Thank you for submitting your manuscript for consideration by Collabra: Psychology. I have now received two reviews of your (and your colleagues’) paper “Cognitive Control and the Implicit Association Test: A Replication of Siegel, Dougherty, and Huber (2012)”. Prior to sending the paper out for review, I had read the paper myself and I have re-read the paper after seeing the reviews. I very much concur with the central treatise of the reviews – that this was a high quality replication of the Siegel, Dougherty, and Huber paper (study 3). Both of the reviewers feel that the paper is already in very good shape. I very much agree with this assessment. As such, this revise and resubmit could best be conceptualized as an accept pending revisions.

In your revision, please highlight the changes made (I would appreciate you submitting a version with change tracking enabled as well as a clean version of the manuscript). In your coverletter, please detail your responses to all of the points raised about your manuscript summarizing the changes when they are made and if changes are not made, please justify that decision. Given the minor nature of the changes, I do not anticipate sending this out for review a 2nd time.

Sincerely,

John E. Edlund

**Reviewer 1**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

This is a competent and interesting replication study and I would like to see it published after some minor revision. The reported analyses and procedures are technically sound and the paper is clear and well-written.

As a co-author on the Siegel paper, my recollection of those studies was that our initial submission sought to document that cognitive control is an important factor in explaining the magnitude of IAT scores. This is what we reported in Experiment 1, using a battery of cognitive control measures and structural equation modeling. We struggled to publish this result, in part because we were scooped (most prominently by Karl Christoph Klauer). In revising Siegel’s study, we identified a potential novel contribution by going a step farther by manipulating the role played by cognitive control. In revising this manuscript, I suggest that you make clear this distinction between the question of whether cognitive control plays a role in the IAT (Experiment 1 of the Siegel study supports this, as well as other studies that do or do not use Stroop as a measure of cognitive control) versus the question of whether the size of the role that cognitive control plays can be manipulated. This distinction is touched on in the discussion, but it should be made clear throughout the manuscript. For instance, the conclusion of the paper currently reads “We were unable to replicate results demonstrating an effect of cognitive control on IAT scores”. If this was a failure to replicate Experiment 1 of the Siegel study, this might be an appropriate main conclusion. But instead, this is a failure to replicate Experiment 3. So the conclusion should be more like “We were unable to replicate the finding that feedback about racial prejudice affects the relationship between the IAT and the Stroop effect. Furthermore, we did not find any reliable relationship between the IAT and the Stroop effect regardless of feedback”.

Personally, I am not surprised that the feedback manipulation effect failed to replicate. The manipulation effects reported in Experiments 2 and 3 of the Siegel paper were weak effects. I am more surprised, however, that the relationship between the IAT and the Stroop effect failed to replicate, given other studies documenting cognitive control contamination of IAT scores. In the 2010 Klauer study on cognitive control and the IAT, they did not use Stroop as one of the cognitive control tasks. The Ito et al. study included IAT and Stroop, although I don’t think they reported the raw correlation between these two measures – instead, they only report the results of latent factor modeling (if it cannot be found, perhaps it’s worth contacting Ito to find out if there was a reliable correlation). Have others examined IAT and Stroop? I just did a 1-minute literature search and dug up a Psychonomics poster reporting a correlation between IAT and Stroop.

In any case, this study provides compelling evidence that Siegel’s Experiment 3 Siegel does not replicate. But why? Was the original study a false positive? (this seems likely). Alternatively, has something changed in the last 10 years? Of note, in the low feedback condition, Siegel failed to find a relationship between the IAT and Stroop, similar to what was found in general in this study, regardless of feedback. Perhaps worry over one’s IAT score has changed (e.g., maybe participants are now more uniformly motivated to do well on the IAT, thus reducing individual differences in response inhibition as an important factor). Finally, perhaps this just reflects the messiness of behavioral research (e.g., perhaps there is a real effect that is very small in magnitude, and more often than not, replications fail to reject the null hypothesis).

Signed. David Huber

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

**Reviewer 2**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

Overall, I enjoyed reading this manuscript and have a positive impression of it. The studies seem well conducted and the analyses appropriate for the study designs. I appreciate that the authors took time to conduct replication studies. My suggestions below are relatively minor.

It would be good if the authors specifically reported consistencies and deviations from the original study. They do at time, but there were a few instances where I was unsure if they followed the same procedures. For example, as best as I can tell, the authors told people they had high or low prejudice, but the original study told their participants they had high or low racial prejudice. I don’t think this difference impacted the results, but it should be noted when there are deviations from the original study.

I would like to see the exact feedback the participants were given. I wasn’t able to find the specific feedback when I looked on the OSF page.

I appreciated the table with descriptive stats for both studies (Table 1), but it might make more sense to give descriptive stats per condition.

It would be nice to see Cohen’s d for all the t-tests.

I don’t think the inferential stats were reported for the small telescopes approach for the condition X AMP interaction for Study 1a.

In the abstract and in the paper, the authors conclude that “researchers interested in manipulating cognitive control should use different manipulations.” I agree with the statement, but I don’t think their data speak to the effectiveness of the manipulation. They found that the manipulation did not influence people’s scores on the IAT. That isn’t the same as finding that the manipulation isn’t effective. It could be that the manipulation was successful, but that cognitive control doesn’t influence performance on the IAT.

In the abstract, the authors reported that they reanalyzed the original data. I’m assuming they are referring to the meta-analysis. I would not consider a meta-analysis that includes new studies to be a reanalysis of the original data.

As I said earlier, my comments were relatively minor and overall, I had a positive view of the manuscript.

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

**Author Response**

Aug 6, 2021

Reviewer #1:

This is a competent and interesting replication study and I would like to see it published after some minor revision. The reported analyses and procedures are technically sound and the paper is clear and well-written.

As a co-author on the Siegel paper, my recollection of those studies was that our initial submission sought to document that cognitive control is an important factor in explaining the magnitude of IAT scores. This is what we reported in Experiment 1, using a battery of cognitive control measures and structural equation modeling. We struggled to publish this result, in part because we were scooped (most prominently by Karl Christoph Klauer). In revising Siegel’s study, we identified a potential novel contribution by going a step farther by manipulating the role played by cognitive control.

    Response: Thank you for this background context.

In revising this manuscript, I suggest that you make clear this distinction between the question of whether cognitive control plays a role in the IAT (Experiment 1 of the Siegel study supports this, as well as other studies that do or do not use Stroop as a measure of cognitive control) versus the question of whether the size of the role that cognitive control plays can be manipulated. This distinction is touched on in the discussion, but it should be made clear throughout the manuscript. For instance, the conclusion of the paper currently reads “We were unable to replicate results demonstrating an effect of cognitive control on IAT scores”. If this was a failure to replicate Experiment 1 of the Siegel study, this might be an appropriate main conclusion. But instead, this is a failure to replicate Experiment 3. So the conclusion should be more like “We were unable to replicate the finding that feedback about racial prejudice affects the relationship between the IAT and the Stroop effect. Furthermore, we did not find any reliable relationship between the IAT and the Stroop effect regardless of feedback”.

Response: We agree with Dr. Huber that our conclusions were too broad. We now have revised this sentence to more narrowly read “We were unable to replicate results demonstrating that the effect of cognitive control on IAT scores can be manipulated.”

Personally, I am not surprised that the feedback manipulation effect failed to replicate. The manipulation effects reported in Experiments 2 and 3 of the Siegel paper were weak effects. I am more surprised, however, that the relationship between the IAT and the Stroop effect failed to replicate, given other studies documenting cognitive control contamination of IAT scores. In the 2010 Klauer study on cognitive control and the IAT, they did not use Stroop as one of the cognitive control tasks. The Ito et al. study included IAT and Stroop, although I don’t think they reported the raw correlation between these two measures – instead, they only report the results of latent factor modeling (if it cannot be found, perhaps it’s worth contacting Ito to find out if there was a reliable correlation). Have others examined IAT and Stroop? I just did a 1-minute literature search and dug up a Psychonomics poster reporting a correlation between IAT and Stroop.

Response: We are aware of two studies that reported correlations between the race IAT and stroop task. Richeson and Shelton (2003) reported significant correlations between a race IAT and the Stroop task. However, Klauer, Schmitz, Teige-Mocigemba, and Voss (2010) reported no significant correlation between flower, race, or abstract IATs. As we raise in the discussion, we agree with Ito et al. (2015) that only certain facets of cognitive control (i.e., task-switching) may predict IAT scores relative to other facets captured by the Stroop task (i.e., general cognitive control).

In any case, this study provides compelling evidence that Siegel’s Experiment 3 Siegel does not replicate. But why? Was the original study a false positive? (this seems likely). Alternatively, has something changed in the last 10 years? Of note, in the low feedback condition, Siegel failed to find a relationship between the IAT and Stroop, similar to what was found in general in this study, regardless of feedback. Perhaps worry over one’s IAT score has changed (e.g., maybe participants are now more uniformly motivated to do well on the IAT, thus reducing individual differences in response inhibition as an important factor). Finally, perhaps this just reflects the messiness of behavioral research (e.g., perhaps there is a real effect that is very small in magnitude, and more often than not, replications fail to reject the null hypothesis).

Response: We agree that given the results of the meta-analysis and small telescopes analysis that the original results seem likely to have been a false positive. However, we also agree that we cannot rule out the possibility that concerns over IAT scores may have changed in the decade since the original study was published.

Signed. David Huber

Reviewer #2:

Overall, I enjoyed reading this manuscript and have a positive impression of it. The studies seem well conducted and the analyses appropriate for the study designs. I appreciate that the authors took time to conduct replication studies. My suggestions below are relatively minor.

It would be good if the authors specifically reported consistencies and deviations from the original study. They do at time, but there were a few instances where I was unsure if they followed the same procedures. For example, as best as I can tell, the authors told people they had high or low prejudice, but the original study told their participants they had high or low racial prejudice. I don’t think this difference impacted the results, but it should be noted when there are deviations from the original study.

I would like to see the exact feedback the participants were given. I wasn’t able to find the specific feedback when I looked on the OSF page.

Response: We now include a quote of the feedback given to participants in the procedure section. The instructions mentioned racial prejudice in particular. For example, in the high-prejudice feedback condition the full instruction was “Your results from this task indicate that you have high levels of racial prejudice.”

I appreciated the table with descriptive stats for both studies (Table 1), but it might make more sense to give descriptive stats per condition.

Response: We have now updated Table 1 to provide descriptive statistics per condition and study.

It would be nice to see Cohen’s d for all the t-tests.

Response: We have chosen to use standardized beta as an effect size because our main comparisons of interest are interactions, which do not lend themselves to the Cohen’s d metric. To be consistent with these main analyses, we provide beta values throughout. However, because we have included the descriptive statistics for each condition at the reviewer’s request, interested readers will be able to calculate the Cohen’s d for any condition comparison of interest.

I don’t think the inferential stats were reported for the small telescopes approach for the condition X AMP interaction for Study 1a.

Response: The reviewer is correct. This is because Study 1a was a successful replication of Seigel Study 3 (i.e., the p-value for the interaction was significant). In this case there is no additional information gained by doing the small telescopes analysis, as the replication cannot be an informative or uninformative failure.

In the abstract and in the paper, the authors conclude that “researchers interested in manipulating cognitive control should use different manipulations.” I agree with the statement, but I don’t think their data speak to the effectiveness of the manipulation. They found that the manipulation did not influence people’s scores on the IAT. That isn’t the same as finding that the manipulation isn’t effective. It could be that the manipulation was successful, but that cognitive control doesn’t influence performance on the IAT.

Response: We have removed this sentence from the abstract. We agree with the author that our manipulation may have impacted participants’ behavior, just not their performance on the IAT.

In the abstract, the authors reported that they reanalyzed the original data. I’m assuming they are referring to the meta-analysis. I would not consider a meta-analysis that includes new studies to be a reanalysis of the original data.

Response: We have removed this sentence from the abstract. We did conduct a reanalysis of the original data, but as this is included in the pre-registration and not the main paper, we agree with the reviewer that it would not be appropriate to mention in the abstract.

As I said earlier, my comments were relatively minor and overall, I had a positive view of the manuscript.

    Response: We thank the reviewer for their time and expertise.

**Editor Final Decision—Accept**

July 25, 2021

Dear Alison Robey,

I have now had a chance to read over your manuscript “Cognitive Control and the Implicit Association Test: A Replication of Siegel, Dougherty, and Huber (2012)”, along with the letter describing the changes you made. Thank you for your responsiveness to the concerns that the reviewers and I raised. I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work, I think it will make an important contribution to the literature and I look forward to seeing it published! I hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point. Our managing editor will contact you in case there are any pre-prodution file related questions. You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely,
John Edlund