

Reply to PCIRR-S2 decision letter:
Smith et al. (2002) replication and extension

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response as well as a tally of all the changes that were made in the manuscript. For an easier overview of all the changes made, we also provide a summary of changes.

Please note that the editor's and reviewers' comments are in bold with our reply underneath in normal script.

A track-changes comparison of the previous submission and the revised submission can be found on: <https://draftable.com/compare/WjnIICwMDxFV>

A track-changes manuscript is provided with the file:

“PCIRR-S2-RNR-Smith-et-al-2002-rep-ext-manuscript-v2-G-track-changes.docx”

Response to Editor: Prof. Chris Chambers

One of the original Stage 1 reviewers returned to evaluate your Stage 2 submission, and I have decided that we can proceed on the basis of this review and my own reading of the manuscript.

The reviewer is broadly positive about your submission while also noting two areas needing attention: whether the exploratory analyses are completely justified, and whether the conclusions are as focused as they should be on the core outcomes. I am interested to see your response to these concerns, but in my own reading, I think addressing the second point may well neutralise the first without needing major changes. I will consider your response/revision at desk, and provided you are able to respond thoroughly, full acceptance should be forthcoming without requiring further in-depth review.

Thank you for the review obtained, your feedback, and the invitation to revise and resubmit.

Response to Reviewer #2: Dr./Prof. Uriel Haran

Are the data able to test the hypotheses?

Certainly. The sample size is large enough, the manipulation check seems to suggest the manipulation worked and everything else is consistent with the original study the current one is replicating.

Thank you for the positive feedback.

Are the introduction, rationale and stated hypotheses the same as in Stage 1?

Yes, with the exception of the addition of a passage discussing the difference between shame and guilt in focusing on the self vs. on one's behavior. This discussion was suggested in the review of Stage 1, and I think it improves the paper.

Have the authors adhered precisely to the registered study procedures?

As far as I can tell, yes. There are many exploratory analyses, but they are clearly labeled as such.

Much appreciated.

Are the unregistered exploratory analyses justified, methodologically sound, and informative?

Not all of them are, in my opinion. For example, the comparison of correlations between shame/guilt reactions and explicit shame/guilt is something I would have moved to an appendix, and in any case would not devote three pages to in the main text. Same for the scenario interactions on pp. 35-36. These results do not teach us anything substantive about the hypotheses or the replicated study.

Thank you for the feedback and suggestion. We gave this serious consideration.

We acknowledge that we devoted much space to the exploratory analyses, and can see how that might make the manuscript a bit harder to digest.

However, given the failure to replicate the target's core findings, we believe that providing the readers with fine-grained analyses aimed at better understanding our findings is much needed here and should be communicated in the main manuscript. Follow-up analyses help clarify important aspects of the findings, and help increase the credibility our conclusions. Some exploratory analyses that we thought were less important, we have already reported in the supplementary materials.

For example, in the regression analyses, we found that when entered simultaneously, explicit shame seemed to be a stronger predictor than guilt for (supposedly) shame-related reactions, which cannot be inferred and is not clear based only on the pre-registered correlation analyses.

For the analyses on emotion-related actions in different scenarios, we only provided a summary in the main text, in hope that they will show that again exposure and moral belief influenced both shame and guilt-related reactions in a similar way. The inclusion of the scenario-wise analyses helps show that the mis-calibration of the manipulation check in the plagiarize scenario is likely not the cause of the current results.

Therefore, in the revised manuscript, we only moved the summary of exploratory analyses originally on pp. 35-36 to the end of section *Shame-related and Guilt-related Reactions* without other major edits.

Are the authors' conclusions justified given the evidence?

Yes. But the conclusions need to be presented more clearly. The key analysis of the replication is the interaction between exposure and emotion. Failing to replicate the result of this analysis from the original paper makes for a failed replication of the entire research. All the numerous other results did replicate but they are not material to the current test. Therefore, the key analysis should be given special emphasis in the presentation of the results and the discussion, so that it stands out from the other replicated findings. Currently, the paper reads more like a series of successful replications (and many, many exploratory analyses), followed by a conclusion that the replication attempt failed, which may surprise less attentive readers.

Thank you for the important comment.

Indeed, the results section includes many tests, making it harder for readers to see the link between results and our conclusion. In the conclusion section, we did not do a good enough job in emphasize the core findings and the reason why this is a failed replication.

In the revised results sections, we now emphasize the inconsistent results of the key tests and added the conclusion of a failed replication at the end of the *Comparing Replication to Original Findings* section.

Examples:

“Crucially, we found no support for the two-way interaction between exposure and emotion, $F(2, 1266) = 1.08, p = .34, \eta_p^2 = .002, 95\% \text{ CI } [.00, .02]$, suggesting that exposure did not have differential effects on shame and guilt. This result is again inconsistent with the original article’s finding that exposure had a greater effect on shame

than guilt. To this point, both key tests failed to find support for the target article's hypothesis." (in the "Exploratory: Scenario interactions" subsection)

"Taken together, we conclude that the current study failed to replicate the core findings of Smith et al., (2002)." (in the "Comparing Replication to Original Findings" subsection)

We also revised the conclusion accordingly to emphasize the key results:

"In this close replication of Smith et al. (2002) with a larger more diverse sample, we found that exposure and moral belief manipulations impacted both shame and guilt similarly. More specifically, we failed to find support that exposure had a greater effect on shame than guilt. We, therefore, conclude this as a failed replication not in support of the distinction made between shame and guilt in the target article (i.e., exposure). In addition to the key confirmatory analyses above, we also found that the supposed shame or guilt-related measures were not uniquely associated with shame or guilt, respectively. We note several limitations that may have impacted our replication, yet not in a way that would explain these differences. Finally, we raised issues regarding the exposure manipulation employed in the target article. We suggest caution in future studies assuming that the distinction between guilt and shame lies in exposure and using the exposure manipulation." (in the "Conclusion" section)