

Reply to PCIRR Stage 2 decision letter #444:
McCullough et al. (1997) replication and extensions

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/MsFfMcIHoNIB>

A track-changes manuscript is provided with the file:

“PCIRR-S2-RNR-McCullough-et-al-1997-replication-extension-main-manuscript-trackchanges.docx”
<https://osf.io/u7cwv>

[Sideonote: We note that we are not familiar with the titles and ranks of the reviewers, and looking for that information proves tricky. To try and err on the side of caution, we refer to all reviewers with the rank Dr./Prof. . We apologize for any possible misalignments and are happy to amend that in future correspondence.]

Reply to Editor: Dr./Prof. Chris Chambers

Thank you for your Stage 2 submission, which was evaluated by two of the reviewers from Stage 1. Both are positive about the completed study. In James Bartlett's review you will find a range of constructive comments and suggestions for clarifying specific points, ensuring consistent reporting style, and reporting additional robustness checks (which he has already taken a look at, thanks to your open data). I hope you will find these comments helpful and look forward to receiving a final revision and response in due course.

Thank you very much for the reviews obtained, your feedback, and the invitation to revise and resubmit.

Reply to Reviewer #1: Dr./Prof. James Bartlett

Overview

The authors submitted a complete stage 2 manuscript which is entirely consistent with their plans at stage 1. There are no substantive changes between the stage 1 and stage 2 manuscript apart from entering the new values. Its refreshing to see replication results that are incredibly like the target study. It makes for a short discussion, but this is why there should be more replication efforts across areas of psychology. I like the feature of commenting on the prediction poll in the discussion and encouraging replication as a positive initiative.

I've split my comments into separate sections below on the manuscript, supplementary material, and reproducibility of OSF files. I have also labelled my comments as to address or suggestions. To address are key points to respond to while suggestions are more stylistic that you can ignore.

Thank you very much for taking the time to review the manuscript again. We really appreciate the positive encouraging opening note and your feedback.

Manuscript

(suggestion) – Abstract – For the extension results towards the end, if there is space, it might be worth commenting on the key findings rather than the overall ANOVA. For instance, how forgiveness and apology was highest in the high empathy group.

Thank you for your suggestion. To improve, we elaborated the contrast effects:

“Extending the replication by manipulating empathy, we found that empathy had a causal impact on forgiveness ($\eta^2_p = 0.08$, 90% CI [0.05, 0.11]) and perceived apology ($\eta^2_p = .02$, 90% CI [0.01, 0.04]). Individuals in the high empathy condition were higher than those in the control and low empathy conditions in forgiveness ($d = 0.60-0.62$) and perceived apology ($d = 0.28-0.30$), with no support for differences the control and low empathy conditions.”

(suggestion) – page 22 – I recommended reporting omega for internal consistency but one thing I did not notice in the stage 1 revision was just using the symbols. Given alpha and omega might not be obvious to all readers, it would be worth specifying in the first instance it represents Cronbach’s alpha and McDonald’s omega, then shortening to just the symbols beyond.

Thank you for raising that issue.

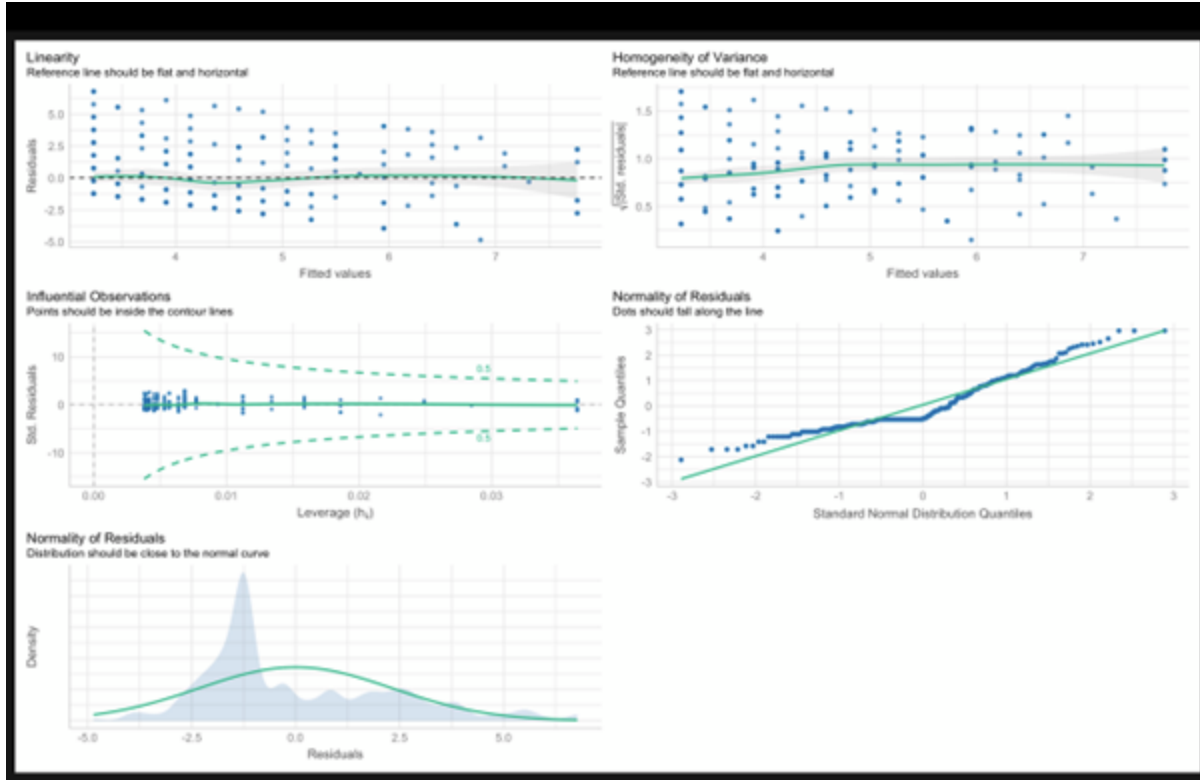
We added the full name of Cronbach’s alpha and McDonald’s omega when first mentioning them, with a blanket indicating their short form. We copied the relevant section below in the subsection of “perceived degree of apology” for your reference.

“We measured the extent to which participants perceived that the offender apologized for the offense with a scale consisting of two items (1 = Strongly disagree; 5 = Strongly agree; scores ranging from 2 to 10; **Cronbach’s alpha** (α) = .85, **McDonald’s omega** (ω) = .85).)

(to address) – Results and page 48 - The issue of following the target article’s analysis techniques regardless of appropriateness came up in the stage 1 report but its something I still think is worth reporting. In the discussion on page 48, the authors conclude there was no impact on the results as the findings were similar to the target article, but this is the wrong comparison as both could be equally wrong.

The good news is the results do seem to be robust to potential problems. I touch on the OSF materials below, but thanks to the authors sharing their data and code, I could quickly explore an example. For example, the relationship between apology and empathy looks problematic in figure 2 since there are many low values in both variables. However, using the performance package’s check_model, it only looks like normality could be an issue (screenshot below). If you switch the correlation method to Spearman, the results are robust, with the correlation changing from .45 to .43.

Long story short, this is the comparison which should reassure the reader that the results are robust. It does not need to change the whole results section, just a note on how the results are robust to potential assumption problems while also honouring the target article’s analysis approach, and the authors can provide further details in the appendix given its more of an exploratory check.



Thank you for checking those and for sharing the results of the analyses. We appreciate that very much. Yes, the associations are strong and appear robust. We added the checks and Spearman rho to the Rmarkdown code, and updated our Table 2 to include Spearman's. We reported this in a new section:

Exploratory analysis: Correlations robustness checks

The variables violated assumptions of normality, and we therefore added Spearman's rho to Table 2 which were very similar to Pearson's correlations and supported the robustness of the associations.

(to address) – page 39 – In the manipulation check, the difference between low empathy and control is described as weaker but not specifically significant or non-significant. It might not be obvious to readers without seeing the output produced in R, but the p-value is literally .05. Its only a manipulation check, but it would be worth specifically commenting on whether you consider it significant or non-significant, since its technically statistically significant precisely on .05.

We appreciate the note. We tried to be very transparent about our results, and have previously reported it as $p = .05$ (which we now tried to make even clearer with $p = .050$). Referring to p-values becomes even more problematic if you take into account multiple comparisons and adjustments, and there was no impact on the contrast between those on the dependent variables. We therefore preferred not to focus on the signal of the contrast between low empathy and control. You might have also noticed that we have opted to try and avoid “significance language” and as much as possible refer to effect sizes and confidence intervals and comparisons of effects.

We would therefore rather keep the framing of the control being close to low empathy than to high empathy.

(Note: We updated our ggstatsplot figures to have nicer labels and also include the pairwise comparisons between the conditions).

(suggestion) – page 40 – once you start reporting the manipulations, the reporting style changes to not include Cohen’s d, whereas it was reported for the manipulation check. Personally, I think it would be useful for readers to see the Cohen’s d reported too, but there is at least the mean difference already.

We appreciate the feedback, and we agree.

We updated the relevant section and added back all the Cohen’s d for our post-hoc Scheffe tests after our ANOVA and manipulations. We copied the relevant paragraphs for your easy reference.

“We conducted post-hoc Scheffe tests and found that forgiveness in the high empathy condition ($n = 264$; $M = 14.9$, $SD = 6.22$) was higher than in the low empathy condition ($n = 267$; $M = 11.1$, $SD = 6.4$; $M_d = 3.80$, 95% CI [2.46, 5.15], $p < .001$; $d = 0.60$, 95% CI [0.43, 0.77]), and than in the control condition ($n = 263$; $M = 10.98$, $SD = 6.38$; $M_d = 3.92$, 95% CI [2.57, 5.28,], $p < .001$; $d = 0.62$, 95% CI [0.45, 0.79]). Yet, we did not find support for differences in forgiveness between the low empathy condition and the control condition ($M_d = 0.12$, 95% CI [-1.23, 1.47], $p = .97$; $d = -0.02$, 95% CI [-0.19, 0.15]).

We conducted post-hoc Scheffe tests and found that perceived apology was higher in the high empathy condition ($M = 5.18$, $SD = 2.94$) than in the low empathy condition ($M = 4.39$, $SD = 2.7$; $M_d = -0.78$, 95% CI [-1.37, -0.20], $p = .005$; $d = 0.28$, 95% CI [0.11, 0.46]) and than in control condition ($M = 4.29$, $SD = 2.61$; $M_d = 0.89$, 95% CI [0.30, 1.47], $p = .001$; $d = 0.32$, 95% CI [0.15, 0.49]). We found no support for differences in perceived apology between the low empathy condition and the control condition ($M_d = 0.10$, 95% CI [-0.48, 0.68], $p = .916$; $d = -0.04$, 95% CI [-0.21, 0.13]).”

(suggestion) – page 41 – for the post-hoc tests, exact p-values are reported everywhere (bar $p < .001$) apart from this paragraph on perceived apology. I would recommend reporting exact p-values here too as a suggestion.

Thank you for the suggestion, we agree, this was an oversight.

We modified the relevant paragraph and reported their exact p-value. We copied the specific paragraph below for your easy reference.

“We conducted post-hoc Scheffe tests and found that perceived apology was higher in the high empathy condition ($M = 5.18$, $SD = 2.94$) than in the low empathy condition ($M = 4.39$, $SD = 2.7$; $M_d = -0.78$, 95% CI [-1.37, -0.20], $p = .005$) and than in control condition ($M = 4.29$, $SD = 2.61$; $M_d = 0.89$, 95% CI [0.30, 1.47], $p = .001$). We found no support for differences in perceived apology between the low empathy condition and the control condition ($M_d = 0.10$, 95% CI [-0.48, 0.68], $p = .916$).”

Supplementary appendix

(to address) – pages 3-5 – In the appendix, the handling outliers and exclusion criteria are still presented as future tense without any comments on the final data. Its presented as “if the replication is unsuccessful”, but it would be at least worth commenting that it was not performed as it was not deemed necessary.

Thank you for catching that.

We modified all the sentences presented in future tense and added an extra sentence to indicate that we didn’t carry out spared analyses thanks to the very successful replication for the target article. We also combined that section with the exclusions section:

“Exclusion criteria and handling outliers strategy

We ran our analyses on the full sample of all those who completed the study successfully and answered all questions. Those who dropped out were not included.

We pre-registered that in the case of a failed replication, as a supplementary analysis and to examine any potential issues, we would also determine further findings reports with exclusions.

[We wrote: “In any case, we would report exclusions in detail with results for the full sample and results following exclusions (in either the manuscript or the supplementary. Criteria:

Participants indicating a low proficiency in English (self-report < 5, on a 1-7 scale)

Participants who self-report not being serious about filling in the survey (self-report < 4, on a 1-5 scale).

Participants who indicated issues or having seen these materials before in the funneling section (manually coded).”]

We also pre-registered that if we fail to find any support for our hypotheses, we would run statistical analyses after excluding the “outliers” with compensation for alpha (i.e. alpha = .01) to account for multiple tests, but this part of our analysis would be purely exploratory since it would deviate from our initial study protocol.

Exclusions and outlier analyses were not carried out given our successful replication of the target article’s findings. We therefore followed the pre-registered plan to include all data in our analyses, without identifying or excluding outliers or applying additional exclusion criteria.”

OSF

I downloaded all the files from the OSF and I could again reproduce all the .Rmd files without edits, so great work here. Its been refreshing to review a paper where the code and data have been reproducible at all steps. The materials are also complete with PDF, Word, and Qualtrics file versions available.

That is wonderful, we hope to see more reviewers going over the materials, data, and code shared in peer-review. Again, thank you very much for your detailed review and constructive feedback. Much appreciated.

Reply to Reviewer #2: Dr./Prof. Saleh Shuqair

Thanks for inviting me to review a replication of McCullough et al 1997 original study.

This article provides a replication of the original study, with similar effect sizes. By doing so, this article not increases the reliability of prior research but goes beyond prior research by manipulating empathy attribution to examine the causality.

Overall, I would like to thank the authors for their efforts. Best of luck in your research

Thank you very much for your review and encouragement.