

Reply to PCIRR S2 decision letter reviews #750: Epley et al. (2008) replication and extensions

We thank the editor and the reviewer for the feedback and below we provide a detailed response to each item. The editor's and reviewer's 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/okISdmiDrtsV> (<https://osf.io/4fq8d>)

A track-changes manuscript is provided with the file:
“PCIRR-S2-RNR-Epley-et-al-rep-ext-manuscript-v2-G-track-changes.docx”
(<https://osf.io/rqc9m>)

Summary of changes

Section	Actions taken in the current manuscript
Discussion	Reviewer 1: <ul style="list-style-type: none">- Added a paragraph to summarize the takeaways- Added a paragraph elaborating our stance on the dilemma of updating original materials

Reply to Editor: Prof. Chris Chambers

I have now obtained an evaluation from one of the reviewers who assessed your Stage 1 submission, and I have decided that we can proceed based on this review and my own assessment. As expected from my own reading of the paper, the review is generally positive and there are few obstacles in the way to final Stage 2 acceptance. Within the comments, you will find some interesting suggestions for clarifying the presentation of results and enhancing the discussion. I look forward to receiving your revision and response, which I will assess at desk before issuing a final recommendation.

Thank you for obtaining the review, providing your feedback, and inviting a resubmission.

Reply to Reviewer #1: Dr. John Protzko

The authors did what they said they would do, so this work should be approved. What I say below are merely suggestions to improve the reporting of the results. I found myself making multiple notes throughout the paper, which were then answered shortly afterwards (happily!).

Thank you for your time evaluating the Stage 2 report and the positive note.

1. My biggest concern is the low anthropomorphism scores in this replication. As far as I can tell, however, the authors did not discuss if it is low compared to the original. What would be best is to present something like a 2x2 grid of density plots of this data (Gadget anthropomorphism, pet anthropomorphism, belief in supernatural, supernatural anthropomorphism) with the mean and 95%CI of the mean indicated, as well as a line of the mean of the original Epley data (where available). That may help elucidate if and how much the scores differ between then and now (I assume the original Epley data is not available).

We appreciate this suggestion.

Unfortunately, Epley et al. (2008) did not report the means of the gadget anthropomorphism measure. They did report the means of the belief in supernatural beings measure and the pet anthropomorphism measure, yet the reported means came from different experimental conditions. As it was not entirely clear how many participants were in each condition, calculating aggregates is tricky. Furthermore, we changed the scale for the belief in supernatural beings measure (from the original 1 to 10 to 0 to 6 here).

Given all these complications, such plots to directly compare our data and the original would not be feasible and/or informative. We added a sentence about this in the Discussion: “though because no descriptive statistics were reported in the original article, we could not assess how different the responses were between the studies”.

2. A minor point is I would like to see more discussion of the results, what do we know now that we did not know before this study was conducted? What insights can be gleaned for future use?

This relates to a bigger discussion on the role of replications, and a failed replication helps audiences update their priors about the existence and the robustness of the phenomenon. We preferred to be cautious and humble about the implications of our findings, and so aimed to keep the Discussion section short.

Thank you for the feedback. We added an additional paragraph at the end to summarize our findings. In general, we believe the most important takeaway from the replication is that it pays to validate measures before using them.

“In summary, our direct replication of Epley et al. (2008) with a large sample failed to find robust evidence for a positive association between chronic loneliness and anthropomorphism. This failure was further complicated by issues such as floor responses and low intercorrelations among different measures originally thought to tap on anthropomorphism. Given the prominence of the three-factor theory (Epley et al., 2007), our findings call for more direct and conceptual replication to establish the link between sociality motivation and anthropomorphism, importantly, with well-validated anthropomorphism measures. Based on our extension, we found a non-trivial positive association between belief in free will and belief in, and anthropomorphism of, supernatural beings. We speculate that this is because people who hold supernatural beings tend not to think of the world as operating under mechanistic, materialistic, and/or reductionist principles, all of which are linked to determinism in philosophical discussions. Future studies can continue exploring the implications of these associations.”

3. The final (and related) point is on the original material used. The authors used the original gadgets. But the results did not replicate. A hostile individual could say “Well, we all know you can’t use the original materials as times change.” But if the authors had used updated materials and still found non-significant results, the hostile individual could similarly say “Well, they changed the materials so it is not a replication.” I know the authors know about this problem. It is a trap, laid by researchers desperate to vilify any nonsignificant replication. I would love to see the authors say this, explicitly, and loudly, in their discussion. There needs to be continued conversation about this ‘updating materials trap’, and this is a good place to continue to point it out.

Thank you for this suggestion. We agree that this point is worth discussing here. Below is what we added to the Discussion:

“While acknowledging that using the original gadgets—but not ones that reflect state-of-the-art technology—might have contributed to the replication failure, we emphasize that replicators often face a dilemma of updating original materials when those materials are assumed to be time-specific (Chandrashekar & Feldman, 2024): if replicators do not update the original study materials and fail to replicate the original finding, evaluators can argue that the materials were outdated and the decision not to update led to the failure; if replicators update the materials and still fail, evaluators can

argue that the update was not appropriate and led to the failure (or that the study is no longer a replication). Both arguments are easy to make, and they can often be legitimate. However, when used to devalue replication studies, they suffer from strong hindsight and outcome biases (i.e., the replicators should have anticipated the consequence of their decisions to update/not update; Chen et al., 2021; successful replications are likely considered trivial regardless of update decisions), are often empirically untestable, and if used on a scale, hinder the development of a literature free of publication bias. We suggest evaluators refrain from evaluating *single* replication studies with these arguments, particularly when the outcomes are already known. Instead, it would be beneficial to take a meta-analytical view and study how these design factors systematically affect the evidence for a claim on a sufficiently large, bias-free literature (Cumming, 2014). This literature should welcome replications of all sorts.”