

## **Reply to decision letter reviews: #181**

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 while our answers are underneath in normal script.

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

**A track-changes manuscript is provided with the file:  
"PCIRR-RNR-Epley et al-rep-ext-manuscript-v4-G-trackchanges.docx"**

Summary of changes

Below we provide a table with a summary of the main changes to the manuscript and our response to the editor and reviewers:

<b>Section</b>	<b>Actions taken in the current manuscript</b>
General	R1: We removed references to the supplementary information in the Abstract.
Introduction	R1: We included a more detailed description of the original studies. We moved the hypotheses to the main manuscript. R2: We cited more studies on the anthropomorphism-sociality motivation link.
Methods	R1: We included a justification and several reference for including attention checks. We also removed a bogus question that was present in our last submission due to the possibility of misinterpretation. We elaborated on the measures (e.g., randomization of scale items, labeling of response options). We removed the trait ranking task. We fixed grammatical errors in the survey. R1/2: Our confirmatory analyses now include only correlations. We have removed the multiple regression models that do not test clear theory-driven predictions.
Results	N/A
Discussion	N/A
Reporting	R1: We will report both analyses with the shorter 3-item scale and the longer full UCLA scale. We added omega as an internal consistency measure and kept alpha, since it is still commonly used.
Supplementary materials	R1: We included one section detailing data collection parameters.

*Note.* Ed = Editor, R1/R2 = Reviewer 1/2

## **Response to Editor: Prof. Chris Chambers**

**I have now received two very detailed and helpful evaluations of your submission. The good news is that both reviewers are broadly positive about the study while also providing constructive suggestions for improvement. Perhaps the main issue to address is clarifying and strengthening both the rationale and implementation of the free will belief extension, to address concerns about its justification and also the underlying precision of the measure. Concerns about this aspect of your design were raised by both reviewers. Beyond this, the reviewers raise a broad range of specific points, from the consideration of literature in the introduction, to the validity of attention checks, to the completeness of reporting (in order to ensure a close replication), as well as a series of technical suggestions (e.g. to eliminate bots from the data). On the basis of these reviews, I am happy to invite a major revision that addresses all points.**

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit. We appreciate your and the reviewers' time and effort.

Disclosure: We would like to note that Reviewer #1, Prof. John Protzko, has worked and co-authored with some of the authors of this project on several collaborative guides and mass collaboration open-science projects.

## **Response to Reviewer #1: Dr./Prof. John Protzko**

**The authors propose to replicate a correlational study, and extend it with some new domains and some new measures.**

**Methodological Review. The authors make a large number of deviations from the original study. Deviations that are important and improve the quality of the study, but deviations nonetheless.**

**Attention check: none were used in the original and no justification is given for their inclusion here. There is a common belief that attention checks improve data quality, but the evidence behind such claims is spotty at best.**

We added a brief justification and several references to the Methods section (under “Attention checks” subsection) about our decision to include attention checks, copied and pasted below:

“Two attention check items were inserted in the UCLA Loneliness Scale and the Free Will Subscale, respectively, each asking participants to select a certain response option on the scale. We excluded participants who failed any of these checks. Adding attention checks was a deviation since none were included in the original study. However, they are low-cost and efficient in helping protect scale validity, identify unserious respondents, and as such, improve data quality (Berinsky et al., 2014; Kung et al., 2018; Shamon & Berning, 2020). As such, we decided to deviate and include them.”

Although these pieces of evidence might seem spotty, as far as we know, there is little clear evidence showing that attention checks can adversely influence the replicability of original findings or psychometric properties of scales. As such, we believe it is better to include rather than exclude attention checks.

On the other hand, we removed the bogus question (“Have you ever been to Mars?”) that was inserted randomly in our survey in our last submission, because it is hard to make sure participants would interpret the question as we do (Curran, 2016, *J. Exp. Soc. Psychol.*; e.g., they may answer “yes” or “not sure” to this question simply because they have dreamt of being on Mars).

**Loneliness scale: the authors are using a longer version that contains the original 3 items. This is nice, but in their results section for only the direct replication of loneliness -> anthropomorphism they should show the results with both the original 3 items and the whole scale.**

Thank you for raising this point. We expanded the Results section (“Loneliness and anthropomorphism” subsection) to also report findings with only the three items from the shorter scale used in the original study. The subsection is copied and pasted below:

“We found a weak positive association between loneliness and anthropomorphic ratings of gadgets,  $r(198) = .02$ , 95% CI  $[-.12, .16]$ , one-sided = .394 (with only the three items of the shorter loneliness scale used in the original study:  $r() = [numbers\ are\ omitted\ because\ these\ are\ similar\ with\ some\ of\ the\ demonstrated\ analyses]$ , 95% CI [], one-sided = []), and a small negative association between loneliness and non-anthropomorphic ratings of gadgets,  $r(198) = -.14$ , 95% CI  $[-.28, -.00]$ , one-sided = .978 (with only the three items of the shorter scale:  $r() = [], 95\% CI [], one-sided = []$ ). Also, we found no evidence that loneliness was positively associated with anthropomorphic ratings of gadgets after non-anthropomorphic ratings were controlled for,  $r_{partial} = .01$ , 95% CI  $[-.13, .15]$ ,  $p = .887$  (with only the three items of the shorter scale:  $r_{partial} = [], 95\% CI [], p = []$ ). Overall, we found little evidence that loneliness was positively associated with anthropomorphism of technological gadgets.”

Note that we do not register analyses with the shorter scale and other dependent measures (e.g., pet anthropomorphism). This is because these analyses are not part of the original study, hence not targeted for replication. We will, however, conduct these analyses in an exploratory fashion, should the analyses concerning gadget anthropomorphism yield different results using the shorter and the longer scale.

**Throughout: there needs to be a little more information on the scales: are the questions for each scale administered in fixed order? Are all the response options labeled or only anchors? Do participants see the numbers on the response options?**

Thank you for raising this point. We added more details to the Methods section concerning the presentation order of scale items and the labels of response options.

“Unless otherwise specified, we randomized the order of items in the measurements that we used; we labeled only the endpoints with text anchors as well as numbers, whereas the middle options only had numeric labels.” (see Design subsection)

For example, for the UCLA Loneliness Scale, since we will label all response options, we now specify them in the text (1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Always). The same is true with other scales/measures.

**Throughout: report omega as a measure of internal consistency instead of Cronbach's alpha.**

Thank you for your advice. We decided to include both alpha and omega, since alpha is still very widely used.

**The authors should add a Captcha at the end, to take into account the bots on MTurk.**

Thank you for the suggestion. We will collect data from CloudResearch approved participants, with a host of measures to address any form of automated answering. Google's reCAPTCHA technology is included in Qualtrics and screened for, which only presents a CAPTCHA if there is some abnormal activity. We believe bots are not extremely unlikely to impossible with all the measures we have implemented.

**The authors should make explicit any requirements (e.g., country, participant completion rate, number of HITs completed, are Masterworkers allowed, etc.) and justify them.**

Thank you for the comment. We documented these requirements in our revised supplementary materials (see "Data collection parameters" subsection).

**The authors randomly order the presentation of the scales, which is another deviation. This is nice and I agree with their decision, but in their results section for only the direct replication of loneliness -> anthropomorphism they should show the results with both the original order (loneliness then gadgets) and then the whole sample.**

Thank you for this suggestion. Our aim is to keep things concise and clear, and presenting multiple analyses that seem very similar would burden readers with very little to no added value. We see including this in the Results as valuable only when there are any indications for differences between the two situations (i.e., loneliness then/after gadgets), or if in the case of a failed replication, and taking order into account would somehow change findings.

Therefore, we will include such analyses as suggested only in these cases, rather than committing to reporting them at Stage 1. We added:

[Note: In case we fail to find support for the original's hypotheses, we will test for order effects (order as a moderator), and results for measures displayed first.]

**I like the extension with free will beliefs, as it is a field I work in and a topic I am interested in. There is a problem, however, in that part of the measure of anthropomorphism includes averaging an item asking if participants believe something or a supernatural object has free will. This creates an uninteresting autocorrelation of people who believe in free will believe other objects have free will. To create a faithful replication, the authors have to include the question, but for analyses involving the belief in free will as a predictor, they will need to create an alternate measure of anthropomorphism that does not include the free will item averaged in.**

Thank you for your valuable input. We respectfully disagree with this comment, as the question of whether people believe humans have free will (what the items from the FWB scale primarily assess) is a very different question from whether they believe various objects have free will, which is at the heart of whether these objects are more likely to be anthropomorphised. This is fundamental to the concept of free will and to the idea of anthropomorphism. It is therefore not an uninteresting autocorrelation.

Also, it is an untested assumption that there is an autocorrelation between whether people believe in free will in objects and free will in people. The literature does not indicate that an autocorrelation between these beliefs exist. In order to provide a more comprehensive picture about free will beliefs, it is important to keep the original analysis but include additional (exploratory) analyses that examine this specific "free will" item in object perception and whether this one differs from other items. This information would not only contribute to the free will literature but also provide the grounding of how to discuss free will and its relationship to anthropomorphism.

However, you have raised an important point, thus we will include a planned discussion subsection regarding the free will extensions and following the findings whether this is an uninteresting autocorrelation or an interesting autocorrelation. This will provide further discussion of what free will beliefs mean in humans and other objects. It is therefore important to be able to assess this assumption in order to move the field forward. We added:

[Based on feedback provided in the peer review process we aim to discuss the free will extension and the possibility of a correlation between free-will beliefs regarding human agents and free-will attributions towards non-humans.]

**Anthropomorphic traits rank: summing the ranks is not a numerical value, it is putting numbers on something masquerading as a number. While I know this is a bigger problem in psychology and it could easily apply to Likert scales as well, it is another bridge to try to draw inferences about the total of a bunch of numbers from a rank order. I recommend the authors cut this measure completely.**

Thank you for the suggestion. We included this measure in our last submission because it was used in another paper by the same group of authors (Epley et al., 2008, *Social Cognition*, Study 1), though they did not sum the ranks but took their average (which isn't essentially different). We found your argument convincing and removed this measure.

**Multiple regression model: I'm not sure the purpose of this model. It does not appear as a test of the 3-factor model, more of a test of what variables of what the authors collected is the stronger predictor. If the authors wish to keep this analysis, I would recommend the following changes: 1) use robust standard errors to account for any data issues; 2) analyze the model as a large path model with all predictors and dependent variables modeled together (this one isn't necessary, but would be an improvement and can provide a very nice graph for readers to understand); 3) test models using a likelihood ratio test as opposed to partial F-tests.**

Thank you for advising on how to improve our analyses. As Reviewer #2 has pointed out, we had a misalignment between our hypotheses/analyses and their justifications. We hypothesized that free will belief can uniquely predict anthropomorphism (which we then tested with multiple regression models), but in the text, we implied that FWB could be linked indirectly to anthropomorphism through a decrease in chronic loneliness. The reviewer went on to suggest that we frame the whole extension as explorative, because we do not have a clear, directional hypothesis about FWB. We decided to follow this advice. Our confirmatory analyses now only include correlations between FWB and anthropomorphism (which would align with the justifications in the introduction) and do not include the multiple regression models.

**Content review. Abstract: reference to supplementary information is not necessary and should be reviewed.**

We removed references to the supplementary information from the abstract.

**p. 3: Given the effect size of the original study, what was the power of the OSC replication (N = 78) to detect that relationship? It should be reported in the MS.**

We should clarify that the OSC replication did not replicate the target original study (i.e., Study 1 in Epley et al., 2008), but Study 3 (and they used one measure from the original Study 2).



According to the power analysis by the replication authors, their sample size had roughly 95% power to detect the original effects.

We do not think it is necessary for us to calculate or report the power of the OSC replication based on the original effect sizes. On the contrary, this could be rather misleading. Given the small N of the original, the obtained effect sizes were very likely overestimates. That the OSC replication had high power to detect these effect sizes does not mean that their sample size was sufficient (and worth referring to) for detecting the true effect sizes.

**p. 6: Somewhere in the introduction before page 6 should be a 1-paragraph overview of the original study (e.g., participants (N = 20) filled out the loneliness scale, were asked about X in gadgets. The original authors showed Y...). The section that starts ‘controllability’ jumps out at the readers who don’t intimately know the study the authors are replicating.**

Thank you for the suggestion. We now provide an overview of the original studies and restructured the introduction to make it easier for readers to follow the manuscript. Our description of the original studies are pasted below:

“In this research, we attempted to replicate Epley, Akalis, et al. (2008). Study 1 of the article measured participants’ anthropomorphism (i.e., the extent to which targets have “a mind of its own,” “intentions,” “consciousness,” etc.) toward a series of technological gadgets, several non-anthropomorphic ratings of the gadgets (e.g., attractiveness), and their chronic loneliness levels (with a short 3-item scale from Hughes et al., 2004). The study found that anthropomorphism and loneliness were positively correlated after non-anthropomorphic ratings were controlled for. Study 2 used an experimental approach and manipulated social connection by suggesting to participants that they would either end up lonely in their lives or be strongly socially connected, apparently based on their responses to a personality questionnaire (Twenge et al., 2001). It was found that those induced to feel socially disconnected had stronger beliefs in supernatural agents (e.g., God, ghosts), regardless of their baseline levels of religiosity. Finally, Study 3 induced socially disconnected, fearful, or neutral (as a control) feelings in participants by having them watch emotionally laden video clips. For dependent measures, participants reported their beliefs in supernatural agents (like in Study 2), chose adjectives that best described their pets from a given list, and noted down what they saw from a series of ambiguous figures. It was found that participants in the social disconnection condition reported higher beliefs in supernatural beings and chose more anthropomorphic adjectives (e.g., thoughtful, considerate) than those in the fear and control conditions. In contrast, participants in the fear condition detected more faces from ambiguous figures than those in the other two conditions. According to the authors, this finding illustrated that the observed anthropomorphism effects were “not simply produced by any negative emotional state,” but specifically by a feeling of loneliness, or social disconnection (Epley, Akalis, et al.,

2008). Overall, their results supported the idea that sociality motivation increases anthropomorphism.”

**p. 10: for the “has consciousness” items, there would be a grammatical error for angels has consciousness (should be have) and ghosts has consciousness (should be have). Will the authors change it?**

Thank you for catching that, much appreciated! We fixed this grammatical error by changing all subjects to their singular forms. As such, participants will be asked if, for example, “an angel” has consciousness.

**p. 12: put the hypotheses in the main manuscript.**

We moved our hypotheses to the main manuscript.

**p. 18: There is no closed bracket on the paragraph about non-convergence.**

We added a closed bracket to the paragraph.

**Overall, I look forward to seeing the author’s response and the eventual study.**

Thank you for your time and feedback. We appreciate it very much.

## **Response to Reviewer #2: Marieke Wieringa**

**The manuscript “Revisiting the positive association between loneliness and anthropomorphism with an extension to belief in free will: Replication and extensions of Epley et al. (2008)” proposes a close replication of the study “Creating social connection through inferential reproduction: Loneliness and perceived agency in gadgets, gods, and greyhounds” by Epley et al., (2008). The study design in the proposed manuscript is based on the correlational study (Study 1) described in Epley et al. (2008), but also includes outcome variables from Study 2 and Study 3 from that same paper. Additionally, it extends the research by adding a potential second predictor (belief in free will) and two additional outcome variables (anthropomorphism of supernatural beings, perceived controllability of gadgets). Overall, the manuscript is very well written, and I believe the proposed method for replicating the target article, while at the same time implementing the proposed extensions is well thought-through and scientifically sound. I do have three recommendations/points of concern for the current version of the manuscript. These include the arguments for replicating this specific article, the free will belief extension, and the planned sample size. Below, I address my thoughts and recommendations in detail. I hope the authors find my recommendations helpful in the development of the manuscript and study design.**

Thank you for the positive opening note. We appreciate your helpful and constructive feedback.

**Arguments for replicating Epley et al. (2008). In the current manuscript, the authors provided several reasons for a new replication of Epley et al. (2008), including the high impact of the original paper and the mixed results of previous replications. The manuscript then mentions two replications in specific (Open Science Collaboration, 2015; Bartz et al., 2016) and highlights the limitations of these replications. I think the argument for replicating the original study can be strengthened by citing a few more empirical studies on the link between loneliness and anthropomorphism (not necessarily replication studies) that yield null or mixed results. Also, I did wonder if another replication is preferable to a meta-analysis of the existing literature on the connection between loneliness and anthropomorphism (given that the original study was cited 700+ times). I recommend the authors to cite more studies here and to briefly elaborate on their choice of a replication vs. a meta-analysis.**

Thank you for the comment. Following your advice, we referred to more studies on the link between anthropomorphism and loneliness in the introduction in order to help readers have a

better overview of the literature (see the “Replication target and motivation for replication” subsection). Nonetheless, we do not believe that “empirical studies on the link between loneliness and anthropomorphism that yield null or mixed results” constitute a precondition for a replication. The literature as of now is highly skewed towards positive results, and findings as high-profile as ego depletion, with hundreds of articles with positive findings, can fail the test of high-powered replications. Our stance is that replications are worthy on their own so long as they are methodologically rigorous, and replicators need not (though they are certainly advised to) consult previous findings to determine if they should start to replicate.

As for whether it is more preferable to conduct a meta-analysis rather than a replication: we do not believe that this is a choice that we have to make here. The community needs both replications and meta-analyses. Meta-analyses without sufficient “open” research (e.g., pre-registered, RRs) might not be informative, and bias correction methods are not panaceas (e.g., Kvarven et al., 2020, *Nat. Hum. Behav.*). Unless it is only us who are tasked to do both, we find it unnecessary for us to justify our prioritizing replications over meta-analyses, as the choice would depend on the resources and expertise the team has.

Kvarven, A., Strømmland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour*, 4(4), 423–434. <https://doi.org/10.1038/s41562-019-0787-z>

**Free will belief extension. I think the rationale behind the extension on “belief in free will” as a predicting factor of anthropomorphism is in need of a little more explanation and support by scientific literature. Specifically, I am not sure whether I am fully convinced by the argument that thinking of free will as a uniquely human trait will “be associated with a subjectively sharper contrast between humans and non-humans and predict weaker anthropomorphism” (p. 5). This is because the very definition of anthropomorphism is to attribute such humanlike characteristics to non-human objects. Other studies on anthropomorphism (Eyssel et al., 2011; Eyssel & Reich, 2013) have used the attribution of traits that are thought to be uniquely human or that reflect human nature (Haslam et al., 2008) as an indicator of anthropomorphism. These studies have reported significant effects of some of the factors described in the Three Factor Theory (Epley et al., 2007) on the attribution of these traits.**

**I therefore wonder whether regarding free will to be a uniquely human trait will, as the authors claim, indeed hinder the attribution of humanlike traits to non-human objects and would like to see more literature in support of this claim.**

We are not entirely sure that we follow the argument, yet what we agree with is that the definition of anthropomorphism entails attributing to nonhumans those characteristics that are *typically considered* unique to humans, and free will may be one of them. It follows that, when one anthropomorphizes, e.g., a non-human animal, for this person (and perhaps at this particular moment), free will is no longer something unique to humans. The reverse could be true, then: if one strongly believes that only humans can have free will, they tend not to attribute it to non-humans (and hence there would be little anthropomorphism). For us, “seeing free will as unique to humans” is not in line with “attributing free will to non-humans” (part of anthropomorphism), as such, we believe our reasoning holds on the conceptual level.

Still, since our main focus is the replication, and we added FWB mainly for exploration, we would rather not go into a lengthy review of the literature here. To the best of our knowledge, there are no studies so far that have explored FWB in association with anthropomorphism - and we believe that this is interesting and of potential contribution as an extension (meant to be a tiny addition to replications that could yield insights that pave the way for future, more comprehensive investigations).

**Additionally, I wondered whether it is possible that belief in free will relates differently to some outcomes than others. Specifically, studies have already shown a positive association between religiosity and belief in free will, as belief in free will is a core concept in some religions (Carey & Paulhus, 2013). A significant correlation between belief in free will and the belief in supernatural beings as measured in the proposed study might therefore also not necessarily be evidence of anthropomorphism. Perhaps the authors could elaborate briefly on this possibility in the manuscript.**

Thank you for raising this point.

Having given this some thought, we hope you'll understand that we came to the conclusion that it goes far beyond the scope of the proposed research to test whether, and to what extent, belief in supernatural beings (and religion) is related to anthropomorphism (or individual anthropomorphic propensity).

There is some evidence that religious belief is correlated positively with anthropomorphism (Waytz et al., 2010, *Perspect. Psychol. Sci.*), but the literature examining these associations also with free will beliefs is rather messy and complex, with mixed findings depending on various factors, and given that this is an exploratory extension, we would rather not go into this in detail. We might be able to discuss whatever we find in the Stage 2 discussion of our findings.

**Finally, the authors hypothesize that belief in free will has unique predictive power. However, the authors also argue that “there is some evidence that belief in free will is positively associated with prosocial behaviors, dispositional gratitude, and importantly, sense of belongingness [...] and as such, negatively associated with the tendency to anthropomorphize” (p. 5/6), which would suggest that the belief in free will might actually be associated with anthropomorphism indirectly, through a decrease in chronic loneliness. I therefore suggest that the authors clarify their claims that belief in free will uniquely predicts anthropomorphism. Alternatively, they could also choose to frame this extension as exploratory work (since the authors also do not have a specific prediction of the direction of the effect), which would mean that they would not formulate in the manuscript that they believe that this variable will have unique predictive power.**

We appreciate your suggestion to reframe the extension as exploratory. We agree that there was a misalignment between our hypotheses and their justification in the previous manuscript, in the sense that we implied an indirect effect that we did not plan to test. Given that we do not have a prediction on the direction of association between FWB and anthropomorphism (if any), we believe it makes the most sense to commit to only exploratory correlational analyses at this stage.

We will surely follow them up should we find anything interesting. Please see also our response to a similar comment by Reviewer #1.

**Oversampling. The authors have conducted a power analysis based on effects reported in previous research on the association between loneliness and anthropomorphism (Bartz et al., 2016). This analysis resulted in a required sample size of 439 participants. The authors however state that they plan on recruiting 1000 participants given their budget and the fact that they would like to be able to detect small samples. I would actually advise against oversampling by that much, as to not unnecessarily burden participants without gaining a lot of predictive power (I have received this exact comment by reviewers in studies where I too oversampled by around 300 participants, based on my available budget). It seems that the power analysis that resulted in a required sample size of 439 participants was already based on a rather small effect ( $f^2 = .029$ ). Of course, some oversampling will be required to compensate for failed attention checks. The authors might want to consider setting their targeted sample size at around  $n = 500$ . With the budget they save, they could run a follow-up study that could corroborate/clarify possible unexpected effects they might encounter in their replication.**

We appreciate the suggestion to reduce the planned sample size, but we hope you will understand our decision not to follow it. Effect sizes in the literature are generally overestimated. Since our budget allows us to collect the number of participants that we originally planned, we prefer to err on the side of caution to ensure that we have sufficient power to detect the actual and potentially small effect sizes and provide a more accurate effect size estimate. A large planned sample size can also help compensate for failed attention checks and make it possible to conduct well-powered moderator analyses (e.g., whether the order of completing the measures has an effect).

A convincing well-powered study here does not limit our ability to run additional studies, but rather quite the contrary - it may help convince funders, original authors, and the community of the findings and that insights gained as are less likely to be due to noise or insufficient power.

We note that we are open to revising this decision if given clear editorial guidelines by the recommender.

**Conclusion. Overall, I think the design of the study is described very clearly in the manuscript, and the proposed methodology of the study is scientifically sound. Though I would ask the authors to pay a little more attention to the rationale behind the free will belief extension, I think the other extensions they propose (anthropomorphism of supernatural agents, perceived controllability of gadgets) make a lot of sense. The concerns I addressed above should be easily addressed in a new version of the manuscript.**

Thank you for your time and feedback. We appreciate it very much.