Reply to PCIRR S1 decision letter reviews #898: Rozin et al. (1997) replication and extension

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response to each item.

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/hUGkahzQKAxb (https://osf.io/pcrzu)

A track-changes manuscript is provided with the file: "PCIRR-S1-RNR-Rozin-etal-1997-replication-manuscript-v2-G-trackchanges.docx" (https://osf.io/vxqch)

[Reminder, in the snapshot PCIRR managerial team wrote: “In relation to the recommender's comment above ("Why is it important to replicate this study?"), note that the importance of the research question (and thus the importance of the replication) is not an evaluated criterion at PCI RR (see criterion 1A here).“]

Reply to Editor: Dr./Prof. Michèle B. Nuijten

> **Thank you for submitting your Stage 1 Report, titled “On the moralization of vegetarianism and differences between health and moral vegetarians: Registered Report replicating and extending Rozin et al. (1997)” to PCI RR. I appreciate the work that went into this submission, especially by simulating and filling out placeholder results, which greatly facilitated the evaluation and comprehension of your proposal.**
>
> **After careful consideration and evaluation by both myself and the reviewers, I believe that your work has potential for an IPA in PCI RR. All three reviewers recognize the merit of your study proposal but have also raised several points that would benefit from further clarification and refinement. I would like to invite you to address these points in a revised submission. I have highlighted below certain aspects that I find of particular importance, along with additional comments of my own.**

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit. We appreciate the thorough review of our work and the help in improving our manuscript. Please see our detailed reply to the reviewers.

> **.1. First, the reviewers agree that the theoretical background should be expanded to provide more context for the research question. I agree that this is important, especially as it may affect the research questions in the extensions. In a related point, the reviewers also ask for more justification why you deem it necessary to replicate the target study, in particular. Please note that “the importance of the research question” is not a criterion that PCI RR assesses. However, I do think it would benefit the paper if you could explain more clearly what your goal is with this replication, as this affects certain decisions and interpretations in the rest of the proposal (as the reviewers also point out).**

Our focus is first and foremost the replication of a seminal finding in the field, and we consider that to be important in and of itself. Given the feedback, we realized that our attempt in adding

extensions and strengthening the target article's design may have shifted focus and confused readers. This is one of the challenges in attempting a replication and extension project. In this revision have made adjustments to make that clear and greatly simplify things. Given that our focus is on the replication and our experience with other PCIRR replication projects, we feel that an in-depth theoretical background or a literature review would go beyond the scope of what we aim to achieve here, and we instead provide enough information for the readers to get the context of the target article for replication to understand what we are attempting.

To address the feedback, we also incorporated some of Reviewer #1's literature suggestions in the introduction and make our aims in this project clearer.

> **.2. Specifically, it matters if your ultimate goal is to draw a conclusion about the trustworthiness of the original findings or if your goal is to draw a conclusion about the general research question. This choice affects (among other things) whether and how you have to define "replication success".**
> **At the moment, you propose to categorize replication success based on the number of confirmed hypotheses. I agree with one of the reviewers that creating an "overall" replication conclusion is not the most informative.**

Our main aim is to revisit and replicate the target article and its hypotheses using its methods and under its criteria, re-addressing their main research question.

When indicating that our choice is "not the most informative" it was not clear whether you had something else in mind you wanted us to implement, and we are unsure what would have been more informative and objective than using an overall replication conclusion. We have previously employed this criteria in many of our PCIRR replications, and these seem straightforward and objective. All hypotheses address a similar research question from different angles, and so prefer a simple objective criteria that encompasses all of them to our subjective criteria of which of the hypotheses should be considered the central ones, and then debating these against others' subjective criteria. Please see our more detailed response on this to the reviewers.

We are open to changing this given clear editorial guidelines on what a better criteria would be, and if possible - would appreciate citations to other PCIRR replications with many hypotheses that have successfully employed the suggested criteria.

> **.3. Another aspect that changes depending on your goal with this replication, is whether it is important to replicate hypothesis 3 (with the reproducibility error). If the goal is to say something about the robustness of the original study, you do not have to replicate hypothesis 3 (as you can already conclude now that it is not robust), but if your goal is to draw conclusions about the theory, there is no reason to discard hypothesis 3 based on reporting errors in the original study. A third aspect that is affected by your goal is the smallest effect size of interest. You now define this as 33% of the effect size found in the target study, but if your goal is to say something about the theory, I would argue you need to define a practically relevant smallest effect size of interest.**

Our goal for this replication is to revisit Rozin et al. (1997) and replicate its hypotheses using its methods and under its criteria. Therefore, to be consistent with Rozin et al. (1997), we aim to revisit all seven of the original hypotheses, regardless of whether the target article found support for them or not.

We revised to make that clearer. We test all hypotheses in the target article, including Hypotheses 3 and 7. However, given that these hypotheses were not supported in the target article, we do not expect to find support for them, and therefore our power estimations do not include these hypotheses and these hypotheses are not part of the replicability success criteria.

We had no claims for broader aims or broader theory, and no aims to define what is a practically relevant smallest effect size of interest for this phenomenon. Those go beyond the scope we set for conducting a replication. We feel that these are subjective and debatable, and we would rather try and stick to clear objective benchmarks given the findings in the target article.

We realized that our inclusion of equivalence testing is what may have led readers to expect a SESOI and shifted focus from our main goals, and we therefore decided to remove the equivalence tests.

We also realized that our inclusion of extensions to try and use the current design to gain additional insights had a similar issue of confusing readers and shifting focus, and so we removed the extensions in the manuscript/supplementary material and have now made our focus more explicit. Please see our reply to the reviewers below on these points.

> **.4. A second point I found of particular importance is the validity and rationale of your measures. One of the reviewers points out that the classification of moral/health vegetarians heavily depends on whether the reasons you classify as moral/health reasons are actually perceived as such. I think this is a crucial point and I wonder if you have/plan to gather evidence that this is indeed the case.**

We appreciate the feedback. Given that this study is a direct replication, we feel it important to run the study under similar terms and to retain as much as possible from Rozin and colleagues' methods, including something as fundamental to their testing as the classification of reasons. That said, we do see the value in validating this classification of moral/health vegetarians to examine whether participants perceive the reasons to be moral or health reasons and whether those match the participants self-identification as either moral or health vegetarians.

To address the first point about the classification of reasons, we now include an additional matrix of questions to query participants' perceptions of whether each reason is moral, health, or neither moral nor health-related. We then added a measure of self-identification as either a moral or health vegetarian. We also added paragraphs of our planned validations in the results section.

> **.5. Furthermore, you decided to expand the range of animals included in the questionnaire, but – as the reviewers point out – this makes it very long which could affect the quality of the answers.**
>
> **Furthermore, the justification of the expansion is not clear to me. What is the main conclusion you would hope to draw from this? My main hesitations here are the specific choice of animals (some of the suggested animals are quite rare and as you are planning to recruit US citizens, I assume that most of these questions will be very hypothetical – e.g., attitudes towards eating gorilla) and the way you aim to statistically combine the information (if I understand correctly, you intend to calculate averages across all animals, in which case you would lose the information on differences in attitudes towards eating common meats such as beef or pork and uncommon meats such as gorilla or mole; which would mainly add a lot of heterogeneity to your data which could negatively affect your power).**

Our main goal is to revisit Rozin et al. (1997) and its hypotheses. We therefore removed our proposed extensions, including the expansion of the range of animals. This makes the survey shorter which we hope would also be less demanding for the participants. We've run far longer

and more demanding studies with our chosen target sample and have repeated and consistently received very high quality data, so we expect that they are up to the task.

> **.6. Third, it sounds to me as if the proposed additional analysis (p. 35) does not match with the research question. The hypothesis is formulated as "there are group differences between moral vegetarians and health vegetarians in the correlations between variables", but the analysis does not make use of the earlier classifications into actual distinct groups of vegetarians but instead uses the full data. The authors follow the target article in calculating the total moral-ecological score and total health score and correlate these with elicitation of disgust, but as I would interpret it, this is not the same as "group differences", or at least a different interpretation of what the groups consist of.**

Thank you for raising this. This seems to be an issue with the target article. The target article's approach of using the total moral-ecological score and total health score and their respective correlations with the variables of interest does not directly test their hypotheses.

This is why we proposed the following additional analysis to more directly test Hypothesis 5:

> "As we discussed in the introduction, the correlations reported in the original article only provided indirect support for Hypothesis 5 as they only realistically demonstrate that higher agreement with moral reasons in vegetarians were associated with higher emotional reactions to eating meat and that this association was descriptively stronger than that between the agreement with health reasons and emotional reactions. Unfortunately, the authors' use of data from the full sample precluded any inference of group differences between moral and health vegetarians. We therefore aimed to conduct a more stringent test of Hypothesis 5 and decided to include a one-sided Welch's t-test of the differences between moral-origin vegetarians and health-origin vegetarians on emotional reactions to eating meat."

Our use of the cocor R package was only meant as a refinement of the original article's approach of descriptively comparing correlations. In our revised data analysis strategy section, we make it clear we do not use the findings of these analyses as a criteria for replication success. Given our focus on replicating all seven hypotheses, we also added the variables relevant to Hypothesis 7 in this analysis.

We revised to the following:

> "In the original study, the authors claimed to have found substantiating support for their hypotheses (Hypotheses 4 to 7; Table 1) – that there are group differences between moral vegetarians and health vegetarians in the correlations between variables – by examining

> the descriptive differences in magnitude between correlations (for example, the correlation between the elicitation of disgust and total moral-ecological score and that between the elicitation of disgust and total health score) and the statistical significance of each correlation. We aimed to improve on their method with formal statistical tests of the differences between correlations (Diedenhofen & Musch, 2015): Using Zou's confidence interval method (Zou, 2007) as implemented in the cocor R package (Diedenhofen & Musch, 2015), we sought to examine whether the differences between correlations with summed moral-ecological score and summed health score for the variables of interest – overall disgust, emotional reactions to eating meat, personality reasons, taste of meat, smell of meat, texture of meat, and appearance of meat – in the full sample are different. We note that these analyses are only meant as a pre-registered refinement of the original article's approach; the findings of these analyses will not be used to test Hypotheses 4, 5, 6, and 7, and are therefore not a criterion for replication success.

> **A small additional point about the statistics: I am not sure how you simulated your results, but I noticed a lot of inconsistent t-test results (p-values that did not match df and test statistics; see https://statcheck.io). If this is due to a simulation artifact – no matter, but if it is due to some type of statistical correction, please make sure to report the fully adjusted result, to maintain internal consistency.**

We understand, thank you for checking.

This could be a result of a number of things like our use of one-sided tests, since when we change our tests to "two.sided" then things match.

In any case, we added the following placeholder as a planned check in our Stage 2 following data collection:

> [For Stage 2: Editor raised possible inconsistencies with stat-check on the statistics using the simulated data, which could be due to the use of one-sided tests or some artifact with the simulated data. In Stage 2, we will check statistical results using the real data against stat-check and work out any inconsistencies.]

Since you inquired, we will just clarify that our simulated data was created using the very simple Qualtrics' "Generate test responses" function that takes our Qualtrics survey and simulates participants that adhere to the survey design.

Reply to Reviewer #1: Dr./Prof. Ben De Groeve

> **Thank you for inviting me to review your planned replication and extension of Rozin et al. (1997), a seminal study on moralization and vegetarianism. I praise the authors for their extensive preparation, and I think they provide a clear and sensible justification for replicating the target article. Given that this article has been published more than 25 years ago, it also makes sense to extend the study. Over the years, the field of vegetarianism has grown substantially, with new studies providing insight on the moralization of animal product comsumption and new measures that have been proposed to measure motivations for meat avoidance. More broadly, there have been interesting developments in the psychology of moralization.**

Thank you for the positive and supportive opening note and the constructive feedback.

> **.1. One study that, to my surprise, was not mentioned in the manuscript is the study of Feinberg et al. (2019), who examined the process of moralization in three longitudinal studies and found that moral emotions (e.g., disgust, guilt) and moral piggybacking (e.g., "“When you think about eating meat, to what extent does it lead you to think about your identity as a moral person?") mediated moralization. I think this study is relevant to mention, as you also examine moral emotions, and the measure of moral piggybacking somewhat overlaps with the "personality reasons" in Table 5; watching animal cruelty videos may cause people to reflect on their moral identity when they think about eating meat.**

Thank you for suggesting further literature. We did not intend to go to an in-depth review of the literature, that is a worthy and complex project on its own, but we appreciate the feedback to try and improve on what we did.

We added a brief discussion of Feinberg et al. (2019), so that the importance of replicating Rozin et al. (1997) is better emphasised:

> “As the first study to examine the consequences of moralization in vegetarianism, Rozin et al. (1997) has also indirectly led to further efforts to understand the process of moralization in the context of vegetarianism and meat eating. Feinberg and colleagues (2019), for example, proposed a model of the moralization process they termed the Push-Pull Moralization Model (PPMM) that explicates how moral emotions and cognitions “push” individuals to moralize whereas hedonic motivations and dissonance reduction strategies have the opposite effect of “pulling” individuals to not moralize.

> Testing this model in the context of meat eating using three longitudinal studies, the authors found evidence to support the notion that moral emotions and engaging in moral piggybacking (associating eating meat with one's existing moral principles) served as "push" factors for the moralization of meat eating."
>
> **.2. More theoretical background**
>
> **I think research on meat-animal dissociation (and compartimentalization) provides additional clues on the moralization process. Most people perceive meat as food, but when people are reminded that meat comes from animals, they experience moral emotions which may motivate meat avoidance (Benningstad & Kunst, 2020). I believe that moral vegetarians moralize meat eating mainly because they view meat not as food, but as remnants of a harmed animal, or more broadly, as a cause of harm. In short, their meat avoidance probably reflects harm avoidance (De Groeve et al., 2022). This aligns with the theory of dyadic morality, according to which moral judgment depends on perceptions of an intentional agent causing damage to a vulnerable patient (Schein & Gray, 2018). When people make the connection between their meat-eating behavior and its harmful consequences, this might cause them to reflect on their self-image as a moral agent (i.e., moral piggybacking) and motivate dietary change. This is also consistent with norm activation theory, in which an awareness of harm and perceived responsibility predict a sense of personal obligation to take action. To better understand the consequences of this moralization process, the paper of Buttlar et al. (2024) seems informative, as they examine meat ambivalence across different stages of behavioral change (also action and maintenance stages). This paragraph is just to share more theoretical background, because I felt this was missing and it might help to interpret findings or further improve the design (e.g., extensions). I will now proceed to more directly comment on the planned study. My main concern is the measurement of reasons for avoiding meat (see below).**

Thank you for the additional resources. Much appreciated.

Our aim with the replication is to focus as much as possible on the target and its very close literature to give readers better context, and so we feel that these references would be a better fit for the eventual Stage 2 discussion, when we attempt to contextualise our Rozin et al. (1997) replication findings in the broader moralization literature.

We added those as a planned discussion for Stage 2:

> [Potential addition of citations to be included in the discussion of our Stage 2 based on feedback from reviewer Dr./Prof. Ben De Groeve: Benningstad and Kunst (2020); Buttlar et al. (2024); De Groeve et al. (2022); Schein and Gray (2018).]
>
> **Abstract**
>
> **.3. I know the findings are simulated, but why do the upper boundaries of the confidence intervals for ds* extend to infinity (∞)?**

Thank you for the feedback. Much appreciated.

We were debating this and initially decided to report one-sided confidence interval because of views arguing that the corresponding 95% confidence interval of a one-sided t-test should be a one-sided confidence interval with an upper/lower bound of ∞ (see the second paragraph of Chapter 7.4 of Laken, 2022; https://lakens.github.io/statistical_inferences/07-CI.html#sec-relatCIp). The aim of doing so is to maintain the direct relationship between the computed *p*-value and the confidence interval. A one-sided 95% confidence interval (with infinity at the upper/lower bound) is also the default output of the t.test function in R when alternative = "greater" or alternative = "lesser".

However, we acknowledge that reporting a one-sided confidence interval can be confusing to readers and introduces challenges in comparing the target's findings to the replication's. To remedy this, we now report two-sided 90% confidence intervals for the Cohen's ds* instead. Doing so maintains the direct relationship between the *p*-value of a one-sided t-test and the effect size CI (https://stats.stackexchange.com/questions/257936/matching-confidence-limits-with-one-sided-hypothesis-tests) and facilitates a quantification of uncertainty around the effect size estimate.

> **.4. Authors use the terms "moral/health vegetarian" and "moral/health-origin vegetarian" interchangeably. Personally, I prefer the former terms for readability, but this is a minor issue.**

Thank you for the feedback, much appreciated.

There is a subtle difference, and we were trying not to use these interchangeably. When referring to moral/health vegetarians, we were referring to those who are in fact moral/health vegetarians, and we were using the terms moral-origin and health-origin vegetarians to align our replication with the original study's methods of identifying vegetarians of the two categories:

> Like in the original study, moral-origin vegetarians were defined as participants who listed moral and/or ecological reasons and no health reasons among the first reasons for

> becoming a vegetarian/vegan. Conversely, health-origin vegetarians were defined as participants who listed health reasons and no moral/ecological reasons among the first reasons for becoming a vegetarian/vegan.

We added the following clarification note:

> We note that in our manuscript when we refer to moral-origin and health-origin vegetarians we are referring to those who - using this method - were classified as either moral or health vegetarians.

We went over the manuscript again to check that all of our references to analyses on moral/health vegetarians using the target's method are referred to as moral/health-origin vegetarians.

We also realized that this can be confusing, and it is possible that someone classified as moral-origin vegetarian might not self-classify as a moral vegetarian and someone classified as a health-origin vegetarian might not self-classify as a health vegetarian, and so we added an exploratory self-report measure:

> <u>Self-identification as a moral or health vegetarian (exploratory)</u>
>
> Participants were asked if they self-identified primarily as a moral vegetarian or as a health vegetarian with the following question: "Do you primarily identify as a vegetarian for health reasons or a vegetarian for moral reasons?" (1 = Vegetarian for health reasons, 2 = Vegetarian for moral reasons).

We also added a validation section at the beginning of our results section to check the alignment between self-identification and the target's method of classification.

Research questions and hypotheses

.5. Given that the authors plan a replication of Rozin et al. (1997), the research questions and hypotheses are based on this study. The authors want to examine consequences of moralization in vegetarians, by comparing moral and health vegetarians' reasons for avoiding meat and their emotional reactions towards eating meat. The authors found that the t values were not calculated correctly in the target article, and found that hypothesis 3 was no longer supported after recalculation. Based on the note of the authors on p.15, I can confirm that this recalculation is correct. Concerning the note of the authors on p.16, I think equivalence testing makes sense in relation to hypothesis 7, though I do not have experience with this type of analysis.

Thank you for helping to check and confirm our calculations.

Extensions

.6. With regards to the proposed extensions (p.17-18), I'm a bit surprised that the authors want to elaborate on hypothesis 3, given that this hypothesis was actually not supported in the target article. Do they expect that the hypothesis will be supported in their study due to a larger sample size?

Thank you. As our goal for this replication is to revisit Rozin et al. (1997) and replicate its hypotheses using its methods and under its criteria, we intend to test Hypothesis 3.

We revised to make sure that this goal is clearer and now state this goal explicitly in a new paragraph, added towards the end of the introduction:

> "We therefore embarked on a Replication Registered Report of Rozin et al. (1997). We aimed to revisit Rozin et al. (1997) to examine the reproducibility and replicability of their seminal findings associated with its seven original hypotheses (see Table 1) by adhering to the original methodology as closely as possible, in an independent pre-registered well-powered close replication. More specifically, we aimed to investigate if we can (i) directly replicate the findings that the original study found support for (Hypothesis 1, 2, 4, 5, and 6) and (ii) if, with a larger sample size, we can find support for the hypotheses that were not supported in the original study (Hypothesis 3 and 7)."

.7. The authors also hypothesize that compared to health-origin vegetarians, moral-origin vegetarians: (1) are more opposed to the use of animals for scientific testing, (2) like a wider range of animals, and (3) reject a wider range of products that directly or indirect involve the use of animals, but the rationale behind these hypotheses is not entirely clear to me. Are these also construed as consequences of moralization?

Thank you for noting this. Given the cumulative feedback and to streamline the current replication study, we removed the proposed extensions to make it clear we focus our efforts on replicating the hypotheses of the original article.

.8. If the authors are interested in consequences of moralization, I think some of the following variables could (also/instead) be considered:
- the willingness to engage in advocacy or support policies (Bryant et al., 2024)
- the rejection of motives to eat meat (Hopwood et al., 2021)
- ambivalent attitudes towards meat (Buttlar et al., 2024).
- perceptions of meat-eaters (e.g., as less moral) or vegetarians (e.g., as more moral, less judgmental) (Aloni et al., 2024)

We decided to remove our proposed extensions and therefore will not be adding any new extensions. We hope that our project will inspire many followup studies to test new directions.

Method: Sample size

The authors conducted power analyses and the sample size is considerably larger than the one in the target article.

Method: Design and procedure

.9. I'm a bit worried about the length of the questionnaire. The authors might consider to shorten it based on pretest results.

Thank you. We removed our proposed extensions, resulting in a much shorter survey.

Method: Measures

Concerning the note on p.27: I agree with the authors that the quotation marks can be removed.

Meat avoidance:

.10. I think this question is mainly interesting if you would ask it before the fixed measures.

The target article did not make use of the open-ended meat avoidance questions in their analyses and we had no plans for those either, so we decided to remove the three open-ended questions to streamline the present study and to further shorten the questionnaire.

Reasons for avoiding meat:

- A downside of this measure is that it is not rigorously validated. For example, the "moral" measure is mainly about animal rights/welfare, but includes one double-barreled item referring to religious beliefs and/or movement membership. The moral label is also confusing, because ecological/environmental motives can also be construed as moral motives (see Rosenfeld, 2019).

.11. - The sum of non-moral, non-health reasons appears to include ecological reasons, but not environmental reasons, which seems inconsistent. Why are environmental and ecological reasons not in the same category?

Thank you for the feedback.

As this is a direct replication of the target article, we aimed to stay as close as possible to how the target article measured the reasons for avoiding meat. Therefore, we did not change the wording of the items or the categorisation of motives. However, we added a planned discussion of this point in our Stage 2:

> [Planned discussion for Stage 2: We plan to discuss challenges and identified weaknesses in the target article's methods. For example, reviewer Dr./Prof. Ben De Groeve noted issues with some of the reasons items as confounding several reasons. Another important example is Dr./Prof. Willem Sleegers's point regarding the classification of vegetarians based on reasons, that may lead to many in the sample not included in the analyses. We included extensions and supplementary analyses to address these points, yet they warrant further discussion and suggested remedies.]

To better align our study with the goal of revisiting Rozin et al. (1997) and replicating its hypotheses using its methods and under its criteria, we decided to remove the two environmental reasons we previously added in the first submission, to follow the methodology used by Rozin et al. (1997) as closely as possible.

> **.12. - To test the hypothesis that moral vegetarians indicate more non-moral(/health) reasons to avoid meat, it seems crucial to ensure that the non-moral reasons are (perceived as) non-moral. For example, the personality reason "Killing and eating animals makes it easier for us to be aggressive and violent" could also be construed as a moral reason (and was indeed highly correlated with "moral/ecological reasons"). Disgust can be a consequence of moral reasons, but not necessarily. In other words, the classification of reasons as "moral" or not by the authors/Rozin et al. (1997) may unintentionally influence whether the hypothesis is supported. Ideally, the authors should pretest reasons for their perceived morality.**

Thank you for raising this point.

Given that this study is a direct replication, we chose to retain Rozin and colleagues' classification of reasons. That said, we agree that the classification of moral/health vegetarians may be contingent on whether participants perceive the reasons are perceived to be moral or health reasons. This is one of the weaknesses in the target article's methods.

To address this point, we included an additional matrix of questions to query participants' perceptions of whether each reason is moral, health, or neither moral nor health-related. We added the following:

> **Perceptions of reasons for avoiding meat**
>
> To determine how participants perceived each of the 20 possible reasons for avoiding meat, participants were instructed to classify each reason using a three-point response scale (1 = moral/ecological reason, 2 = health reason, 3 = non-moral/ecological and non-health reason).

As an added safeguard in the event that Rozin et al. (1997)'s method of classifying participants as moral versus health vegetarians fails, we will also include the following question to ask participants if they self-identify as a moral or a health vegetarian:

> **Self-identification as a moral or health vegetarian (exploratory)**
>
> Participants were asked if they self-identified primarily as a moral vegetarian or as a health vegetarian with the following question: "Do you primarily identify as a vegetarian

> for health reasons or a vegetarian for moral reasons?" (1 = Vegetarian for health reasons, 2 = Vegetarian for moral reasons).

We added a paragraph to detail that we pre-registered the following:

> As the current replication attempt is a direct replication of Rozin et al. (1997), we chose to retain the author's method of classifying participants as moral or health vegetarians – using self-reported time of onset on a pre-categorized set of reasons for vegetarianism. We will validate Rozin et al. (1997)'s method of classifying participants as moral or health vegetarians, we will check the alignment between participants' self-identification and the categorization of health/moral-origin vegetarians by Rozin et al. (1997). We also pre-register that if we fail to validate Rozin et al. (1997)'s method of classifying participants as moral or health vegetarian, or if more than 10% of the sample will not be moral/health-origin vegetarians according to their method, then we will also report the analyses using participants' self-identification as a moral or health vegetarian to test the seven original hypotheses. In that case, replication success will still be determined based on the analyses using the target article's original criteria, yet in the case of inconsistent conclusions based on the two analyses our replication conclusion will be noted as "with reservations" with added details about the inconsistency and likely cause.

We also added validations of the target article's classification of the reasons and participants' type of vegetarians.

We also added a planned discussion of these and other weaknesses in our Stage 2 discussion.

> **- The authors correctly state that new reasons for vegetarianism have emerged over the years. Recently, Stahlmann et al. (2024) published a new measure of vegetarian motives (VEMI+), including a measure of health, environment, animal rights, disgust, social, pandemic and zoonotic diseases, and farm workers' rights motives.**
>
> **.13. - If you want to adapt the current measurement of reasons or use new measures for reasons to test the hypotheses 1 and 2, you can also find non-moral motives for food consumption in the measure library of PHAIR (p.16-17). Although I understand that a direct replication requires little deviation in the original design to enable comparability, I think deviations can be justified if they allow to better test hypotheses.**

Thank you for the helpful resources.

To better align our study with the goal of revisiting Rozin et al. (1997) and replicating its hypotheses using its methods and under its criteria, we decided to remove the two environmental

reasons we added, to follow the methodology used by Rozin et al. (1997) as closely as possible. Therefore, we also refrain from adding new reasons to the original list of reasons.

We added the following to the discussion section:

> [Planned discussion for Stage 2: We originally had several suggestions for extensions, yet given reviewer feedback decided to focus our replication on the replication. We will briefly suggest the extensions we removed as future directions and potentially add reviewers' suggestions offered in the peer review process as improvements to the removed extensions.]

> **.14. - Concerning: "In the original study, participants were instructed to interpret the term meat as beef, unless they do not avoid eating beef."**
>
> **If the the latter is the case, the participant would not be a (dietary) vegetarian, so this formulation appears a bit strange to me.**

We agree that the original wording of this instruction is puzzling, yet we kept it to be consistent with the target article.

We added a note in our planned Stage 2 discussion to signal that such oddities will be discussed in the eventual manuscript:

> [Planned discussion for Stage 2: We tried to follow the target article's method as closely as possible, yet we note several decisions by the authors that we found puzzling, which we will outline and discuss in Stage 2 as potential improvements for future research. Example: participants were instructed to interpret the term meat as beef, unless they do not avoid eating beef.]

> **Extensions:**
>
> **.15. See above. I'm not sure about the theoretical value of these extensions, in particular concerning the 45 animal meats (e.g., gorilla, lion, elephant) and the liking of animals. In my opinion, this takes a lot of space in the questionnaire which could be used for other extensions, but I think it's up to the authors to decide given that they already prepared a lot of work.**

To focus on our main goal of revisiting Rozin et al. (1997) and its hypotheses and to minimize participant burden, we decided to remove our proposed extensions.

Data analysis strategy

I agree that it makes sense to conduct one-sided Welsh's t-tests.

.16. With regards to the extensions, do you plan to do conduct t-tests for each type of animal meat separately, or will they be grouped? How is the "range" calculated?

To simplify all that, we removed our proposed extensions. In the replication we mirror that target article's analyses.

.17. Outliers

It might still be useful to check for the impact of outliers.

Thank you for the suggestion. Our experience with outlier removal in replication work is that this creates more challenges than solutions, especially given that the target article did not employ any such techniques. Beyond the challenge of comparing a replication with outliers to a target article without outliers, there are heated debates about how to detect and treat outliers, especially in the context of replications when the replication interpretation changes based on the outliers. The use of outliers can also have a major impact on the power to detect the effects in the original study.

We used fixed scales throughout (rather than open range responses), so we consider the potential impact of outliers to be limited to none. We therefore decided to not check for the impact of outliers.

(Some of many recent example citations showing the challenges:

- Checking outliers with *performance*. (2025, March 22). Retrieved from https://easystats.github.io/performance/articles/check_outliers [has a good discussion of the many challenges in how to employ outlier analyses.]
- Miller, J. (2023). Outlier exclusion procedures for reaction time analysis: The cures are generally worse than the disease. *Journal of Experimental Psychology: General, 152*(11), 3189–3217. https://doi.org/10.1037/xge0001450
- Sullivan, J. H., Warkentin, M., & Wallace, L. (2021). So many ways for assessing outliers: What really works and does it matter? *Journal of Business Research*, 132, 530–543. https://doi.org/10.1016/j.jbusres.2021.03.066)

.18. Replication - equivalence test (p.41)

I have no experience with this test, but I'm a bit confused about the null hypothesis of "no effect", because I read a null hypothesis in equivalence tests does not assume "no effect"? (For example, see https://en.wikipedia.org/wiki/Equivalence_test).

Thank you for the feedback. We are unsure, but this may have reflected a misunderstanding.

In any case, to streamline our analyses and to avoid confusing the readers about what our goals for this replication are, we removed the equivalence tests.

Supplementary

The authors provide a clear overview of comparisons and deviations between the target article and the planned replication.

.19. Overall, I appreciate the thoroughness and preparation evident in this replication study, and I am confident it will offer valuable insights into the consequences of moralization and vegetarianism upon revision.

Thank you for the review and the positive and constructive comments.

Reply to Reviewer #2: Dr./Prof. Willem Sleegers

> **The authors propose to replicate Rozin et al. (1997). I think the authors do a great job in handling many aspects of conducting a replication study. Below I offer some of my thoughts, grouped under Major comments, Minor comments, and Writing comments. Naturally, I think my Major comments are the most important and warrant serious consideration. In short, I think there's much to like about the paper, although I think some of the bigger picture elements of a replication can be improved, such as why the target study needs to be replicated and what the goal is of the replication study. Hopefully my comments below will help in improving the paper.**

Thank you for the positive and supportive opening note and the constructive feedback.

> **Major comments**
>
> **.1.1. I think more could be done to justify why Rozin et al. should be replicated. One of the arguments given is that it has been cited 731 times. This is a decent number of times, but it's not so much that I think it's obvious this is a very influential paper. The authors also write that "We believe that Makel et al.'s (2012) note arguing "if a publication is cited 100 times, we think it would be strange if no attempt at replication had been conducted and published" captures the importance of revisiting such impactful findings", but I'm not sure how much value to place on this argument. It reads more like an opinion that replications should be conducted more (which I agree with), rather than an argument about the value of this specific study.**

Thank you for sharing your views.

The PCIRR managing team already noted the following on our snapshot:

> "[...] note that the importance of the research question (and thus the importance of the replication) is not an evaluated criterion at PCI RR (see criterion 1A here)."

Addressing the broad view of the value of replications, some of our views on this topic are shared in:

> Feldman, G. (2025). The value of replications goes beyond replicability and is associated with the value of the research it replicates: Commentary on Isager et al. (2024). *Meta Psychology*. https://doi.org/10.17605/OSF.IO/BTNUJ

Summarized with the following quote:

> Replications are still often misunderstood and undervalued. Despite ongoing discussions regarding the importance of replications, there has been little to no progress in making replications mainstream. One of the strongest indicators is the replication-nonreplication publication ratio, with recent estimates putting the ratio of replications of publications at around 0.2% in psychology (Clarke et al., 2023) and 0.54% in education (Cook et al., 2024), unfortunately closely resembling the rates reported a decade ago with 0.1% in psychology (Makel et al., 2012) and 0.13% in education (Makel et al., 2014). Accumulating meta-scientific evidence indicates replications are still an anecdote, at best, with most research not subject to independent direct replications. [...]
>
> At the moment, given how scarce replications are, almost any replication of published articles that were not yet replicated is of value. We need to conduct, submit, and publish more replications. [...] It would take years if not decades and a mindset shift to come anywhere close to a novel-replication publication ratio that makes sense for credible science.

In this manuscript, we preferred to cite a seminal paper in our field on their views rather than referring to our own. We felt that this note captures widely held sentiments, yet we understand if some scholars do not share these views. We feel that this is an important debate to engage in, but that this debate goes far beyond the scope of this project.

Regarding the specific point about 731 citations as not indicative of being an influential paper, based on some estimates of the citation distributions in social psychology (see Haslam et al., 2008), a paper having 731 citations is an extreme outlier. Even if one disagrees with the view that it is strange that an article with 100 citations has not had a direct replications, we mention the citation rates for this article and that quote hoping that atleast readers would agree that it would be even stranger if there was a seminal article with 731 citations from the 1990s with no direct independent well-powered replications. We do not feel like using citations as a measure for replication value is needed, but if for some it is, then an example citation regarding the use of citations as a baseline measure for “influence”:

- Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J., & Wilson, S. (2008). What makes an article influential? Predicting impact in social and personality psychology. *Scientometrics*, 76(1), 169–185. https://doi.org/10.1007/s11192-007-1892-8
    - Quote: “The 308 articles were cited an average 39.94 times (SD = 41.21), with a very wide range (0 – 338).”
- Isager, P. M., van Aert, R., Bahník, Š., Brandt, M. J., DeSoto, K. A., Giner-Sorolla, R., ... & Lakens, D. (2023). Deciding what to replicate: A decision model for replication study selection under resource and knowledge constraints. *Psychological Methods*, *28*(2), 438.

- Isager, P. M., van't Veer, A., & Lakens, D. (2021). Replication value as a function of citation impact and sample size.

.1.2. The next argument from the authors is that the findings of Rozin et al. (1997) were the foundational empirical evidence for Rozin's (1999) paper, but it's not explained how. The process of moralization is not explained, nor how the findings of Rozin et al. (1997) relate to that process. It's not even clear to me how Rozin et al., (1997) can speak to a process of moralization, since the study appears to be wholly cross-sectional, so how can it reveal something that appears to be temporal (i.e., a process).

Thank you for the feedback. We were hoping to keep this very concise and focus on the empirical aspects of the replication, yet we appreciate the nudge to make the context clearer.

We revised to the following:

> "Beyond the academic attention it received as measured by citation count, the findings of Rozin et al. (1997) are also important because they served as foundational empirical evidence for the consequences of moralization in Rozin's (1999) subsequent treatise on the process of moralization. Rozin (1999) provided the first exposition of the process of moralization, laying out one of the most widely used definitions of moralization to date (Rhee et al., 2019). Therefore, by providing empirical support for the consequences of moralization, Rozin et al. (1997) effectively laid the groundwork for many important follow-up studies that explicated the role of moralization in a myriad of domains including but not limited to health behaviors (Pratt et al., 2024), COVID-19 reduction (Graso et al., 2021), intrinsic motivation at work (Kwon et al., 2023; Kwon & Sonday, 2024), and health products/medications (Lalumera, 2023)."

.1.3. The third argument, that Rozin et al. (1997) sparked interest among other researchers, is also rather weak unless the subsequent work depends on the work by Rozin et al., but there's no persuasive case being made that the subsequent work really hinges on Rozin et al.

Thank you for pointing this out. We have slightly reworded this section to make it clearer that Rozin et al. (1997) itself is not directly responsible for this subsequent work. That said, given its role as the first study to examine the consequences of moralization in vegetarianism, its role in driving this line of research cannot be understated:

> "As the first study to examine the consequences of moralization in vegetarianism, Rozin et al. (1997) also sparked increased research interest in the purported consequences of moralization in vegetarians, particularly in the emotion of disgust. Fessler et al. (2003), for example, sought to extend upon the findings of Rozin et al. (1997) by investigating if

> disgust was a consequence or cause of the moralization of vegetarianism. Similarly, Rothgerber (2014) sought to examine the differences between strict vegetarians and semi-vegetarians on an array of attitudes toward meat and animals (including disgust). More recently, Ioannidou et al. (2023) examined differences in moral emotions (disgust and guilt) and beliefs between omnivores, pescatarians, vegetarians, and vegans.
>
> Rozin et al. (1997) has also indirectly led to further efforts to understand the process of moralization in the context of vegetarianism and meat eating. Feinberg and colleagues (2019), for example, proposed a model of the moralization process they termed the Push-Pull Moralization Model (PPMM) that explicates how moral emotions and cognitions "push" individuals to moralize whereas hedonic motivations and dissonance reduction strategies have the opposite effect of "pulling" individuals to not moralize. Testing this model in the context of meat eating using three longitudinal studies, the authors found evidence to support the notion that moral emotions and engaging in moral piggybacking (associating eating meat with one's existing moral principles) served as "push" factors for the moralization of meat eating."

> **.1.4. I suppose I am missing more of a discussion of the moralization theory. It would be nice to discuss this, how the hypotheses derive from it, and how the results support them or not. For example, to me it does not necessarily strike me as necessary that a moral vegetarian has more reasons for being vegetarian than a non-moral vegetarian. In fact the opposite could even be true because having a single powerful moral reason is likely enough to shape behavior while more non-moral (potentially less powerful) reasons might be needed otherwise. A brief discussion of the theory and ways the hypotheses might or might not derive from it could be another argument that justifies replicating the study.**

Thank you for the feedback. We moved and adapted a paragraph from the "Rozin et al. (1997): Summary of hypotheses and findings" section up to the main introduction, so it is clearer right from the beginning how the original predictions are derived:

> "In this article, the authors sought to investigate the consequences of moralization in vegetarianism by studying the differences between moral- and health-origin vegetarians on a set of attitudes and emotional reactions. They reasoned that if meat avoidance has indeed become a moralized issue in moral vegetarians, they should tend to display stronger attitudes, emotional reactions towards meat, and more dislike of the sensory qualities of meat than health vegetarians where meat avoidance has ostensibly not been moralized. Further, they argue that moral vegetarians, particularly those who have been moral vegetarians for a long time, are likely to give more reasons for meat avoidance than

> health vegetarians, because selective information processing (driven by initial moral reasons for avoiding meat) is likely to lead to an accretion of motives."

> **2. I don't really follow the reasoning to not focus on Hypothesis 3. According to a re-calculation of the reported statistics, the difference between moral vegetarians and non-moral vegetarians is no longer significant on the outcome measure related to hypothesis 3, and this is the reason for not including it. One of the other hypotheses also didn't show a statistically significant difference, but that one is not excluded, so I don't follow the logic here. It seems to me like the theory should dictate what the relevant hypotheses are and whether they are worth testing or not.**

Thank you for the feedback. Please see our reply to the editor above.

In summary, we revised to make more explicit our goal of revisiting Rozin et al. (1997) and to replicate its hypotheses using its methods and under its criteria. We therefore test all seven of the original hypotheses, regardless of whether the original study found support for them (and regardless of what the authors claimed). We do not include the unsupported hypotheses in our power analyses or our replication success criteria.

> **3. I am missing a clear discussion of the power and meaning of the effect sizes of the original study. I know that the Cohen d's of the original study are discussed on p. 16 and p. 17, but it seems valuable to me to discuss these also in the context of what the original study was capable of finding. Was their statistical power high enough to find meaningful effects? Are the effect sizes similar to other effect sizes from other papers on the same topic? Or social psychological in general? What does the size of the confidence intervals of the various effects tell us? Relatedly, I think the sample size of the target study is not even mentioned at this point in the paper, which seems would be good to know.**

Thank you for the feedback.

We now mention the sample size of the original more explicitly in the second paragraph of the section:

> "In the target article, the authors recruited 119 participants for their study, of which 36 were classified as moral-origin vegetarians and 26 were classified as health-origin vegetarians."

We also included a sentence in the same paragraph to better contextualise the statistical power of the original article:

> "A sensitivity power analysis using these reported values (assuming alpha = .05 and 80% power) suggests that the original study had 80% power to detect effects as small as d_s = 0.65 (which according to some benchmarks can be regarded as a large effect in social psychology; Jané et al., 2024)."

> **4. I am not really convinced that the extensions discussed on p. 17 are necessary. It's true that "Moral-origin vegetarians and health-origin vegetarians may also differ on other attitudes towards animals", but I'm not sure what the added value is compared to the measures that are being included. I also went through the survey myself and I found it was quite the list of questions, so I could also see that the study might be improved by making it shorter.**

We appreciate the feedback. To streamline the study and to focus on our primary goal of reproducing Rozin et al. (1997), we removed our proposed extensions.

> **5. Maybe relatedly, I don't get a clear sense of what the goal of the replication is. For example, is it the goal to see whether the hypotheses are correct but that the effect sizes might be smaller compared to the original? Or is it the goal to show that the statistical evidence of the original is unlikely/not compelling?**

We now state explicitly our goal to revisit Rozin et al. (1997) and replicate its hypotheses using its methods and under its criteria. That is, if we adhere to the target article's methodology in our replication, are we able to find support for their hypotheses?

We wrote the following:

> "We therefore embarked on a Replication Registered Report of Rozin et al. (1997). We aimed to revisit Rozin et al. (1997) to examine the reproducibility and replicability of their seminal findings associated with its seven original hypotheses (see Table 1) by adhering to the original methodology as closely as possible, in an independent pre-registered well-powered close replication. More specifically, we aimed to investigate if we can (i) directly replicate the findings that the original study found support for (Hypothesis 1, 2, 4, 5, and 6) and (ii) if, with a larger sample size, we can find support for the hypotheses that were not supported in the original study (Hypothesis 3 and 7)."

6. The authors include an assumption that 10% of the sample might not be classified as either a moral vegetarian or non-moral vegetarian. Is there any pilot testing of this to see whether this is a plausible number?

Thank you. We meant this as an estimation, not as an assumption, and changed our framing accordingly.

Given feedback from the other reviewers, we now also include a self-classification measure that forces the participants into indicating themselves to be one of the two, which would allow us to conduct supplementary analyses based on the entire sample.

We added the following to our data strategy in the results subsections about validation:

> As the current replication attempt is a direct replication of Rozin et al. (1997), we chose to retain the author's method of classifying participants as moral or health vegetarians – using self-reported time of onset on a pre-categorized set of reasons for vegetarianism. We will validate Rozin et al. (1997)'s method of classifying participants as moral or health vegetarians, we will check the alignment between participants' self-identification and the categorization of health/moral-origin vegetarians by Rozin et al. (1997). We also pre-register that if we fail to validate Rozin et al. (1997)'s method of classifying participants as moral or health vegetarian, or if more than 10% of the sample will not be moral/health-origin vegetarians according to their method, then we will also report the analyses using participants' self-identification as a moral or health vegetarian to test the seven original hypotheses. In that case, replication success will still be determined based on the analyses using the target article's original criteria, yet in the case of inconsistent conclusions based on the two analyses our replication conclusion will be noted as "with reservations" with added details about the inconsistency and likely cause.

We also added a planned discussion of this point in our Stage 2:

> [Planned discussion for Stage 2: We plan to discuss challenges and identified weaknesses in the target article's methods. For example, reviewer Dr./Prof. Ben De Groeve noted issues with some of the reasons items as confounding several reasons. Another important example is Dr./Prof. Willem Sleegers's point regarding the classification of vegetarians based on reasons, that may lead to many in the sample not included in the analyses. We included extensions and supplementary analyses to address these points, yet they warrant further discussion and suggested remedies.]

> **7. The authors write that "The current replication attempt will be considered a successful replication if of the five hypotheses (Hypotheses 1, 2, 4, 5, and 6), four or five were supported, a mixed replication if two or three were supported, and a failed replication if one or none of the hypotheses were supported." It seems wholly unnecessary to me to classify the success of the replication in this way. It seems quite arbitrary and ignores whether some of the hypotheses might be more important than others, in addition to it not being necessary. The relevant question is whether the hypotheses are replicated or not, not the study as a whole.**

We appreciate you sharing your thoughts on this.

In our experience, we found that some readers do care about article level conclusions. What we are trying to avoid is an imposition of our subjective value judgments of what is and what is not important to replicate by giving a more objective clear criteria. By predefining categories of replication success at Stage 1, we hope to address this group of readers and reduce our flexibility in our and others' determining this post hoc after looking at the results. We feel that a generalized criteria based on the number of hypotheses supported is a far better approach than our subjective value criteria for the hypotheses and prefer to treat them as equal.

If relevant, it might be useful to note that we previously successfully implemented similar strategies in many of our other completed PCIRR replications, and this seemed to have worked out well. For example:

> Chan, M. & Feldman, G. (2025). Factors impacting effective altruism: Revisiting heuristics and biases in charity in a replication and extensions Registered Report of Baron and Szymanska (2011). *Royal Society Open Science*. https://doi.org/10.17605/OSF.IO/BEP78
>
> Endorsed by Peer Community in Registered Reports.
> [PCIRR Stage 2 endorsement/Open peer review]
> [PCIRR Stage 1 recommendation/Open peer review]
> [Preprint] [Open materials/data/code] [Open Access]
>
> Zhu, Z., & Feldman, G. (2025). Revisiting the Psychology of Waste: Replication Registered Report of Arkes (1996). *Royal Society Open Science*. https://doi.org/10.17605/OSF.IO/GF8RC
>
> Endorsed by Peer Community in Registered Reports.
> [PCIRR Stage 2 endorsement/Open peer review]
> [PCIRR Stage 1 recommendation/Open peer review]
> [Preprint] [Open materials/data/code] [Open Access]

Wong, C., & Feldman, G. (2025). Choice Bracketing revisited: Replication and extensions Registered Report of seven experiments reviewed in Read et al. (1999). *Royal Society Open Science*. https://doi.org/10.17605/OSF.IO/VDQEK
[Stage 2 official acceptance/Open peer review]
[Stage 1 in-principle acceptance/Open peer review]
[Preprint] [Open materials/data/code] [Open Access]

8. The authors write that "we determine the consistency of replication and original effects by examining if a given replication effect is in the same direction as its corresponding original effect, instead of examining if the 95% confidence interval of said replication effect include the point estimate of its corresponding original effect." but I'm an argument justifying this.

Thank you for the feedback. We now implement this in the revision.

We initially deviated from LeBel et al. (2019) because we were reporting one-sided 95% confidence intervals and it did not make sense to see if the original effect sizes lay between the lower bound of the confidence intervals and infinity.

We revised our manuscript to report two-sided 90% intervals now, so we now write the following:

> "In line with the interpretative framework proposed by LeBel et al. (2019), we determined the consistency of replication and original effects by examining if the two-sided 90% confidence interval of the replication effect size estimate included the point estimate of its corresponding original effect size estimate (see the "Replication versus the original" section of the supplementary material)."

Minor comments

1. Part of the research question is tautological. The question: "Do moral and health vegetarians differ in their reasons for being vegetarian" is by definition true since moral and health vegetarians are defined by having different reasons for being vegetarian.

Thank you for this feedback. We agree this requires further specification.

We amended the research question to the following:

> "Do moral and health vegetarians differ in the number of reasons for being vegetarian, their attitudes towards vegetarianism and eating meat, and their emotional reactions towards eating meat?"

> **2. The authors mention on p. 15 that there's a discrepancy between the degrees of freedom but don't elaborate on it. Did all of the tests have this discrepancy? If so, what are some of the possible reasons for it? How big were the discrepancies?**

Thank you for raising this issue. We amended this paragraph specify which tests were found to have discrepant findings, to supplement the more detailed analysis in the power analysis supplementary material:

> "Using data collected from these moral-origin and health-origin vegetarians, the authors argued to have found support for six out of their seven hypotheses (Hypotheses 1, 2, 3, 4, 5, and 6). While analyzing the target article for this replication attempt, we noticed that the degrees of freedom (df) for all the t-tests except the t-test associated with Hypothesis 4 (see the "Power Analysis" section of the supplementary material for more details) reported in the target article were discrepant from the df values that are to be expected. Therefore, we recomputed the t-tests using the summary statistics reported in the target article and reported the recalculated t-values, degrees of freedom, and p-values below instead (see "Power Analysis" in the supplementary materials for more details). After our recalculations, Hypothesis 3 was no longer supported (see Table 1)."

> **3. On p. 16 and p. 17 the Cohen d effect sizes have confidence intervals but the correlations do not. I know the authors write that the original study does not mention the sample size, but you could make some explicit assumptions and include them anyway. Also, I know the authors write that "we were not able to accurately compute CIs for the reported correlations", but you can still try and give some estimates rather than implying it can't be done.**

Thank you.

We revised Table 3 accordingly to add confidence intervals based on the lowest of the possible range and added the following in the table note:

> The authors of the target article stated that the sample sizes used for these correlational analyses ranged from 95 to 104. Based on a reviewer's request, we calculated 95% CIs based on the smallest of the range n = 95 (with the help of code in Jane et al., 2024).

4. I don't know what the 'safeguard power approach proposed by Perugini et al. (2014).' is. It would be better to briefly describe the method (which doesn't seem that difficult to describe based on my brief reading of Perugini et al.)

Thank you for this suggestion.

We added a brief description of the safeguard power approach in the power analysis section of the manuscript:

> "The safeguard power approach (Perugini et al., 2014) is a conservative method , aiming to protect against an overestimation of true effect sizes. It suggests that the effect size replications should aim their power analyses to detect the lower bound of target article's effect size 60% confidence intervals."

5. Removing the quotation marks from 'meat' in the instructions seems fine to me.

6. The authors write that "We felt that this list was too limited and that there was a need to distinguish between attitudes towards the consumption of animal meats and that towards animal products/by-products.". I think the phrasing is a bit too informal ('we felt') and unclear (why is it needed?).

Given the feedback from the editor and the other reviewers, we decided to focus on replicating the original hypotheses, and will no longer add extensions.We removed this section.

Is there a benefit of using the cocor package over simply running a model that includes the interaction test, in order to show whether the correlations differ by group?

Given that the present study is positioned as a direct replication, the proposed analysis (using the cocor package) was chosen to match the analysis performed in the original article as closely as possible. Please see our reply to the editor on this point.

> **7. The authors write on p. 37 that "note that because they are all "individual tests of individual null hypotheses" and not tests of an intersection null hypothesis (Rubin, 2024, p.3) we did not adjust alpha for multiple comparisons". I'm not familiar with Rubin (2024) but I can easily imagine this claim could be controversial. It's not at all clear to me why multiple comparisons are not a potential problem here. In fact, if you say that they are all individual tests, then to me that seems like a reason why multiple comparisons IS a problem.**

To streamline our analyses and to focus on our replication goal of testing the seven original hypotheses, we removed our proposed extensions. We now simply follow the target's method and analyses.

> **Writing comments**
>
> **I found the structure of the text to make it a bit difficult to understand everything in the right order. Several times things were discussed or mentioned without the necessary details being discussed first. For example, in the abstract Rozin et al. is mentioned but is not briefly summarized to help readers unfamiliar with the study understand what the core measures and findings were.**

Thank you.

We made the following changes to the abstract, to briefly summarize the goal of Rozin et al. (1997). Due to space constraints, we are limited in our ability to elaborate on what the core measures and findings of the original study are in the abstract:

> "Moralization is the process by which formerly morally neutral objects/activities acquire moral qualities. Rozin et al. (1997) proposed that moralization is responsible for the purported attitudinal differences between moral vegetarians and health vegetarians and therefore sought to explore the differences between moral- and health-origin vegetarians to study the consequences of moralization. [...]"

> **Another example is that the reasons for replicating the study are explained before Rozin et al. is actually summarized. It would be useful to first describe the target article is more detail and afterwards discuss its limitations, strengths, and influence, followed by the details related to replicating it.**

We restructured the manuscript such that Rozin et al. (1997) is first summarised before we discuss our reasons for replicating this study.

Reply to Reviewer #3: Dr./Prof. Seth Green

> **Seems like a nice project and I look forward to seeing the results.**

Thank you for the positive and supportive opening note and the constructive feedback.

> **You've basically convinced me that the paper is important (though I don't agree with the claim that every paper with 100+ citations should be replicated -- that does not seem like a good use of our collective time 😃). I was looking forward to hearing why the results are important to you -- is this your subfield? Were its ideas influential on your work in particular? I am someone who accepts the idea that most published research findings are false, so my threshold for replicating a past study is a bit higher. Ideally it would be one that you, or people around you, had special interest in. I think you get part of the way there when establishing its influence, but I didn't really understand why I should care because I didn't get a sense of why *you* care.**

Thank you for sharing your views and asking.

This deserves a much bigger debate that goes far beyond the scope of this manuscript. As an initial explanation, please see our reply to other views made by the other reviewers above (e.g., Reviewer 2 #1.1) where we also noted the following:

> The PCIRR managing team already noted the following on our snapshot:
>
> > "[...] note that the importance of the research question (and thus the importance of the replication) is not an evaluated criterion at PCI RR (see criterion 1A here)."
>
> Addressing the broad view of the value of replications, some of our views on this topic are shared in:
>
> > Feldman, G. (2025). The value of replications goes beyond replicability and is associated with the value of the research it replicates: Commentary on Isager et al. (2024). *Meta Psychology*. https://doi.org/10.17605/OSF.IO/BTNUJ
>
> Summarized with the following quote:
>
> > Replications are still often misunderstood and undervalued. Despite ongoing discussions regarding the importance of replications, there has been little to no progress in making replications mainstream. One of the strongest indicators is the replication-nonreplication publication ratio, with recent estimates putting the ratio of replications of publications at around 0.2% in psychology (Clarke et al., 2023)

> and 0.54% in education (Cook et al., 2024), unfortunately closely resembling the rates reported a decade ago with 0.1% in psychology (Makel et al., 2012) and 0.13% in education (Makel et al., 2014). Accumulating meta-scientific evidence indicates replications are still an anecdote, at best, with most research not subject to independent direct replications. [...]
>
> At the moment, given how scarce replications are, almost any replication of published articles that were not yet replicated is of value. We need to conduct, submit, and publish more replications. [...] It would take years if not decades and a mindset shift to come anywhere close to a novel-replication publication ratio that makes sense for credible science.

Briefly, our view is that for science to be credible, replications should be mainstream. We consider a singular study using small samples in a specific context and point in time with no open-science (no open materials/data/code or peer-review) to be anecdotal, at best, and so it is part of science and an excellent use of our time and resources to validate such findings with an independent well-powered bias-controlled replication (like a Registered Report) adhering to the highest standards of open-science, before we proceed to do anything with these findings. We have no assumptions regarding whether published findings are false or not, we simply do not know and do not have sufficient evidence to conclude that, and so we need high-quality rigorous accumulating evidence and replications.

This project is part of a mass-replication effort where we revisit seminal findings in the social psychology and decision making literatures. Given its impact on the literature, we consider this to be a seminal finding in social psychology, and we think stakeholders would benefit from better and more updated evidence. This is what we wrote and explained very briefly in the manuscript, assuming that the need for replications is taken for granted (as noted by PCIRR managing team). We would rather not go into more details about our personal motivations for doing this replication or our views on replications more broadly in this manuscript.

> **There are some ways in which the writing can be tightened up a bit, but it can probably wait until publication (it's definitely good enough for a manuscript at this stage). For example, when I look at this sentence:**
>
> **> One such attitude is disgust towards meat; several studies have claimed to have found evidence supporting the notion that moral vegetarians are more disgusted by meat than health vegetarians**
>
> **I would go with something like**
>
> **For example, several studies have found evidence that moral vegetarians are more disgusted by meat than health vegetarians are.**
>
> **or "has had much impact on the literature" can become "has been influential." I also see that you use the word effectually on p. 10 where the appropriate word is effectively, delineated on p. 8 where I think distinguish would work better, & In essence, this serves as a test and an extension of the target article' is missing an s at the end of article.**
>
> **(BTW the phrase 'target article' feels a litlte warlike to me but perhaps that is a convention I just don't know 😃)**

Thank you. We appreciate the effort in helping us improve.

We previously used the term “target article” in all of our previous Registered Reports conducted with PCIRR. We mean nothing warlike about it, aside from it being the target of our replication effort. It is just a straightforward, clear way of referring to the chosen article for our replication efforts.

We made the suggested changes to improve the flow of writing:

- Original: “One such attitude is disgust towards meat; several studies have claimed to have found evidence supporting the notion that moral vegetarians are more disgusted by meat than health vegetarians (Hamilton, 2006; Rothgerber, 2014; Rozin et al., 1997).”

  Revised: “Several studies have found evidence supporting that moral vegetarians are more disgusted by meat than health vegetarians (Hamilton, 2006; Rothgerber, 2014; Rozin et al., 1997).”

- Original: “Rozin et al. (1997) has had much impact on the literature”

  Revised: “Rozin et al. (1997) has been influential on the literature”

- Original: “By way of Rozin (1999), Rozin et al. (1997) effectually laid the groundwork for many important follow-up studies that explicated the role of moralization in a myriad of domains including but not limited to health behaviors”

  Revised: “Therefore, by providing empirical support for the consequences of moralization, Rozin et al. (1997) effectively laid the groundwork for many important follow-up empirical studies that explicated the role of moralization in a myriad of domains including but not limited to health behaviors ”

- Original: “Researchers have often delineated between two types of vegetarians (Antonovici & Turliuc, 2020; Dai & Leung, 2024; Fox & Ward, 2008; Hamilton, 2006; Jabs et al., 1998; Rothgerber, 2014)”

  Revised: “Researchers have often distinguished between two types of vegetarians (Antonovici & Turliuc, 2020; Dai & Leung, 2024; Fox & Ward, 2008; Hamilton, 2006; Jabs et al., 1998; Rothgerber, 2014)”

The sentence containing “In essence, this serves as a test and an extension of the target article…” has been removed because of other revisions we made.

We would appreciate further feedback following our Stage 2 submission.

> **A few technical notes. First, you distinguish between reproducibility and replicability, but these aren't universally established terms, so I'd either define them or stick to replicability.**

We added a footnote where these two terms are first mentioned in our manuscript under section “Choice of target article for replication: Rozin et al. (1997)” with a citation of where these terms have been formally defined::

> See Nosek et al. (2022) for definitions of the terms “reproducibility” and “replicability”.

> **Second, you distinguish between two types of cohen's d, please define them.**

We expanded on the footnote in the “Rozin et al. (1997): Summary of hypotheses and findings” section to describe the differences between the two types of Cohen’s d in more detail:

> “...For comprehensiveness, we computed and reported two effect size measures from the results of the recomputed t-tests here: Cohen’s d_s and Cohen’s d_s^* (see Table 2). Cohen’s d_s is computed by dividing the raw sample mean difference of the variable of interest by the pooled error term of the two groups, whereas Cohen’s d_s is computed by dividing the raw sample mean difference by the square root of the non-pooled average of both group’s variance estimates…”

> **Third, you spend a lot of time on the minium detectable effect and power, but I am more interested in: what's a meaningful effect size here? What does it mean for someone to have more reasons for being vegetarian than someone else -- twice as many? five times as many? I really have no idea. what are your beliefs on this question? And here I'm not that interested in a cohne's d metric -- literally can you tell me how many reason count as a litlte or a lot?**

We understand why you would raise that question, yet we feel that this goes beyond the scope of what we aim to do here with this replication which is to revisit the original findings. We do not feel that it is necessary for us to define this for this replication. We would rather not go into the subjective discussion of what effects mean here and what constitutes a meaningful effect size.

In our power and sensitivity analyses, we aimed to give the target article a good chance to replicate well given an objective criteria based on the effects observed in the target article.

Reporting the Cohen's *d* standardised effect sizes is helpful to allow for easier comparison to, aggregation with, and future use in other studies.

> **My final conceptual note: you spend a fair bit of time trying to reproduce their original results. Again, as a reader, I don't really care about whether a t-test on an article from three decades ago is precisely reproducible. That doesn't change my beliefs about the world at all because I assume errors in this category are common and I have no idea what this article's publication process looked like. For my tastes I'd just stick to: we find these questions interesting because...we think a meaningful effect size would be...this previous paper found [whatever] and we found [whatever].**
>
> **But all of thtis can be adfdressed at the publication stage. I look forward to seeing the results!**

We do care, and we find these important.

We aim to meet objective scientific replication criteria, we do not assume errors or lacking reproducibility or replicability, and we find the issue of reproducibility and reliability extremely important and informative. We aim to help ensure that our published science is credible and updated, and a needed step for that is repeating the same design again as closely as possible while adhering to the most up-to-date highest standards to examine whether the results align. We would hope that the findings from our replication would provide those who care about and/or rely on these effects another high-quality updated data point to adjust their beliefs about the phenomenon.