Reply to decision letter reviews: #190

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/GgAzEqWUVLGb

A track-changes manuscript is provided with the file: Bastian-etal-2012-Replication-PCIRR-RNR-Manuscript-v3-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:

Section	Actions taken in the current manuscript
Introduction	R2: Edited the hypotheses order in the PCIRR table and description of original findings to be congruent with Table 1. R3: Changed the language of the opening paragraphs to be less ambiguous and to moderate claims regarding the paradox framing.

Section	Actions taken in the current manuscript
Methods	Ed; R1; R4: Added exploratory analyses to address concerns about order effects caused by combining the two studies. R2: Changed recruitment platform to Prolific and moved the question about meat-eating to the end of study to avoid activating identities. Changed the manipulation check to be voluntary and moved it to after the mental capabilities scale in Study 2. Clarified the nature of the manipulation checks and why we are using them. Changed the attention check to be consistent in terms of timeframe and tense. R3: Removed ambiguous language regarding the sufficiency of power analyses. Removed section on manipulations and removed mentions of “Replication” and “Extension” from headings and subheadings. Removed outlier analysis. Clarified replication success or failure criteria. R2; R3: Clarified the exclusion criteria and that they will only be used in an exploratory fashion. R4: Changed the instructions for Study 1 to let participants know the range of animals and that they are presented in no particular order.
Results	Ed; R1; R4: Added analyses to address concerns about order effects caused by combining the two studies. R2: Added ANOVA analysis examining interaction between animal species and food status.
Supplementary materials	R2: Updated the items for the attention and manipulation checks as described above; added “Future Directions” section with a paragraph on the sheep/lamb description and a paragraph on adding a sample of vegetarians in the future. R3: Added a paragraph to the “Future Directions” regarding the absence of a filler task. Clarified which sections are to be completed after data collection. R2; R3: Clarified the exclusion criteria and that they will only be used in an exploratory fashion. R4: Added a paragraph on the “Future Directions” regarding possible changes in the meat paradox over time.

Note. Ed = Editor, R1/R2/R3/R4 = Reviewer 1/2/3/4

Response to Editor: Prof. Chris Chambers

> **I have now received four very helpful expert reviews of your Stage 1 submission. In general, the reviewers find substantial merit in your proposed replication and the signs are promising for achieving Stage 1 in-principle acceptance (IPA) in due course. There are, nevertheless, a number of significant issues to address to satisfy the Stage 1 criteria…This summary of issues is far from comprehensive, and you will find enclosed in the reviews a variety of other points -- often accompanied by constructive proposals for solutions. On this basis I am happy to invite a Major Revision and hope you will find these reviews as useful as I have found them.**

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit. We responded to the feedback and made revisions to the manuscript, and we believe that the manuscript is much improved as a result.

1. **One of the major concerns raised by the reviewers is the decision to combine study 1 and study 2 into a single survey using a within-subjects design. The reviews query this deviation from the original methodology and the risk of introducing order effects (despite randomisation) that may introduce ceiling effects (in at least a sub-sample) or otherwise hamper replicability and validity. One solution to this problem that you may want to consider would be to double the sample size, running 50% of participants in one order and 50% in the opposite order. You could then analyse the first session in a between-subjects (closer) replication of the original study while also using the total dataset for any within-subjects analyses.**

We thank you and the reviewers for requesting clarification regarding our reasoning for combining the two studies.

Based on our experience with running replications in the judgment and decision-making literature, this design helps address concerns regarding the sample and attentiveness. When we have some failed studies and some successful studies, then in a separate design one may raise concerns that the failed experiments were due to sample/time/context, yet with a single unified design, that concern is addressed with the much more likely explanation that the failed replications are because of the differences between the studies.

In addition to this benefit, the proposed design and sample target in our previous submission of $N = 1000$ allows us to test the hypotheses in a way that considers any potential carry-over effects from one study to the other. We propose to do this by to testing Study 1's predictions in those participants who completed Study 1 first ($n = 500$). Likewise, we also propose to testing Study 2's prediction in those participants who completed Study 2 first ($n = 500$). In this way, we will present a set of estimates that are not affected by any potential carry-over effects from study-to-study.

We note that we were aiming for increasing the chances for us to detect the effects, and the practical impactions for the request to commit to these analyses are that this effectively reduces the power and ability to detect the effects and is a much stricter test of the target's findings than we planned.

However, we do not need to further increase our sample size, because this is what we already did in our initial submission doubling from the calculated 448 to 1000. Half of the present sample (N_{total} = 1000, n_{half} = 500) has sufficient power (95%) to detect the safeguard effect size (r = .33, d = 0.26), given the observed effects in the original article, this is a reasonable test of the hypotheses.

We will also examine correlations between the two studies as an indirect test of whether the outcomes of the combined studies are converging. We have amended the data analysis strategy to document this approach:

> "Next, we will examine Pearson correlations between the Study 1 and Study 2 measures to examine the degree to which the combined studies were associated, with a positive correlation indicating that participants were responding to the studies similarly. Additionally, we will use moderated multiple regression analyses to test if study order moderated the results of Study 1 and a mixed ANOVA to test if study order moderated the results of Study 2. We will also rerun the primary analyses considering only those participants for whom the study was displayed first."

> **The reviewers also raise substantial concerns about beliefs in animal emotion and cognition having changed since the publication of the original study 10 years ago; sampling bias due to the way in which "meat-eaters" are selected in the inclusion criteria (deviating from the original study); the validity and implementation of the manipulation check; and the structural organisation of the manuscript, which at least one reviewer felt didn't provide essential methodological detail in the best place.**

Thank you for summarizing the feedback- we respond to these comments in detail below.

Response to Reviewer #1: Prof. Brock Bastian

> **I have read over everything and it all looks good to me. There is only the concern that I have, and which I think is quite substantial. Why are the authors combining Studies 1 and 2 into a single survey? This seems to be quite problematic to me. If I am thinking about the minds and edibility of animals, and then I am thinking about a sheep or cow in a paddock, won't the concept of eating animals be primed for me already and potentially impact on how I think about that animal? I think it would be a fair criticism to say that this design could undermine a clean experimental manipulation - comparing thinking about animals as food vs. not as food.**
>
> **The other direction, where Study 1 replication comes second would also be impacted. As people who have read about animals being harmed in the meat production process may be motivated in their ratings of edibility and mind.**
>
> **I think this a fair criticism of the suggested approach and would likely undermine the validity of the replication attempt.**

Thank you for taking the time to review our submission and for sharing your thoughts. We take your concern seriously and agree that carry-over priming effects are important to consider. We believe the combined design has benefits (see comment to Editor above), especially in the event that the findings of one study replicate but another does not, so we have tried to find a solution that preserves these benefits whilst also affording a more valid test of the hypotheses that is not confounded by any potential carry-over effects.

We think that committing to testing the effects within those participants who completed each study first presents a solution. That is, irrespective of what we find when looking across the entire sample (n = 1000), we commit to also testing Study 1's predictions in those participants who completed Study 1 first (n = 500). Likewise, we commit to testing Study 2's prediction in those participants who completed Study 2 first (n = 500). Although this does effectively half our sample size for these analyses, we nevertheless still have high power (95%) to detect a substantially smaller effect (r = .33, d = 0.26) within these sub-samples than what was originally found. We hope these newly-proposed analyses alleviate your concerns and that you now consider the proposal a fairer test.

Response to Reviewer #2: Prof. Ben DeGroeve

> **I am pleased to review the Stage 1 Registered Report entitled "Revisiting the motivated denial of mind to animals used for food: Replication and extension of Bastian et al. (2012)" in accordance with the Stage 1 criteria listed in the Guide for Reviewers of the PCI initiative.**

Thank you for your time and thoughtful feedback.

> **The scientific validity of the research question(s). Given the reproducibility crisis in psychological science and the current evidence showing a motivated denial of food animal minds, the proposal to replicate Studies 1 and 2 of Bastian et al. (2012) using highly powered samples is theoretically justified. The research questions are clear and summarized in the PCIRR-Study Design Table on p.5 and are answerable through quantitative research. The scientific validity of the study is covered in the introduction.**

> **2. In the introduction, the authors provide a rationale for replicating the original findings of Bastian et al. (2012). The proposed hypotheses are precise and summarized in the PCIRR-Study Design Table on p.5, though it seems like the H1b (moral concern) and H1c (negative affect) have been switched. In addition, in Table 1 on p.9, H1b refers to negative affect and H1c to moral concern, while in the text on p.9 the findings are mentioned in reverse order. I recommend to improve consistency in presenting/ordering the hypotheses and associated research questions.**

Thank you for catching this inconsistency. We reordered the PCIRR table and the text in the introduction to match the order of Table 1 (H1b = negative affect; H1c = moral concern).

The authors will gather a well-powered sample and provide a clear summary of their study design (Table 3, p.14). Attention is payed to critical design features such as rules for inclusion and exclusion, randomisation and reducing survey fatigue. The authors also did an analysis with simulated data. Overall, the methods seem feasible, but I question the soundness of three design aspects. (1C-1) Contrary to Bastian et al. (2012), the authors plan to exclude vegetarians and vegans in the beginning of the survey with the following item: "We are running a replication of a classic studymeant for those who eat meat. Therefore, this survey is only for those who self-identify as meat eaters. If you are not a meat-eater (e.g., a vegetarian or a vegan), please return the HIT now. Please indicate: Do you eat meat? with options "Yes, I eat meat" and "No I do not eat meat"."

3. However, people rarely explicitly identify as a meat-eater because eating meat is the norm, so if participants read that the survey is only for those who self-identify as meat-eaters, they might not feel it is about them or they might feel pigeonholed. If the survey is meant for those who eat meat, I don't think it is necessary to explicitly refer to the meat-eater identity. More importantly, the item might prompt participants to think that (not) eating meat is an essential aspect of the study and previous research suggests that a mere exposure to vegetarians might arouse meat-related dissonance (Rothgerber, 2020). My concern is that this item might confound/influence the results of the study because, for example, participants may be more likely to see the non-food animal condition in Study 2 as having something to do with the fact that they eat meat and consequently see the non-food animal more like a food animal. In Bastian et al. (2012), vegetarians were identified and excluded at the end of the survey, so participants' dietary identity was less likely to be salient. To avoid this confounding risk, authors could use the same approach as Bastian et al. (2012) and/or use a prescreening tool to select people who eat meat unbeknownst the participants. I know this is possible in Prolific; I don't know about MTurk.

Thank you for your thoughts on having people explicitly identify as meat-eaters.

We agree that a more straightforward approach would be to recruit participants from Prolific and use their dietary filters to recruit only meat-eaters. We no longer mention dietary identities in the recruitment advert, nor do we ask about them in the beginning of the study. We now measure dietary identification at the end to confirm that they are meat-eaters. We have also made edits to the manuscript to reflect the change in platform from mTurk to Prolific.

> **4. The design of Study 2 largely corresponds with the design of Bastian et al. (2012), but the original design is rather complicated. What is the reason for the within-subjects design and for comparing the mental capacity of a non-food sheep with a food cow (or of a non-food cow with a food sheep)? If we accept the within-subjects design, I speculate that participants would probably respond more consistently if they have to assess the mind of the same animal twice (i.e., non-food followed by food condition) so that it would be more difficult to find an effect. This might arguably be circumvented by using different animals, though this rationale, or the rationale for a within-subjects design, is not mentioned in the current manuscript. In other words, I think the experimental set-up of Study 2 requires more justification.**

Thank you for the comment. The reason for the within-subjects design follows from the aim to conduct a direct replication of Bastian et al. (2012). We think that using two different animals, as Bastian et al. (2012) did and we propose to do, is a strength of the paradigm. In Bastian et al. (2012) and in the present proposal, the experiment uses a pair of animals: a lamb and a cow. The first animal participant's judge is randomly selected from the pair and described as a non-food animal. The second is the remaining animal and described as a food animal. Thus, in Bastian et al. (2012) and in the present proposal, participants always judge two different animals. We have tried to better communicate this aspect of the paradigm in the Design section of the Methods.

> "In Study 1, participants were asked to rate each animal's mental capacities (10 items), edibility (2 items), negative affect about eating it (1 item), and how morally wrong it would be to eat it (1 item). They did so for 8 animals that were randomly selected from 32. <u>In Study 2, participants were shown two animals. The first was described as an animal that will be removed to other paddocks (nonfood condition), and the second was described as an animal that will be killed and made into meat product for human beings (food condition). The animals were a cow and a lamb, which were randomly assigned to either the nonfood or food conditions. In other words, if a participant first saw a sheep depicted as a nonfood animal, then the cow will be later depicted as the food animal, vice versa</u>**.** Below each picture, participants were asked to rate the perceived mental capacities of the animal. At the end of the survey, participants were asked to answer some demographic questions. Summary tables and detailed experimental instructions for Studies 1 and 2 procedures are available in the supplementary (see Table S8)."

> **5. Contrary to Bastian et al. (2012), the authors include manipulation checks in Study 2, which helps to ensure whether participants read the scenarios carefully. Nevertheless, I want to caution the authors that the manipulation check of the non-food condition might also affect participants' response to the food condition and make it easier for them to correctly guess the hypothesis related to Study 2. Potential options include pretesting, forcing participants to stay a small amount of time on the pages, avoiding to reveal the food condition in the non-food manipulation check and/or removing/replacing the non-food manipulation check.**

Thank you for this word of caution regarding this manipulation check. To address this concern, we changed the manipulation check to a non-forced manipulation check afterwards the mental capabilities questions in Study 2, and will simply report how many participants answered the manipulation check correctly.

We enlarged, bolded, and underlined the caption to make it more apparent and used italics to draw attention the relevant parts of the manipulation.

> **6. Relatedly, what would happen to participants who fail the manipulation check? In the procedure (p.19), the authors write that participants have to answer correctly to continue. Will participants be informed about this in the consent form?**

Thank you for pointing out this vagueness. We changed the manipulation check so that it will simply be used as a measure of whether participants read and understood the manipulated captions. We will examine the number of failed manipulation checks as measure of attentiveness across the sample. In exploratory analyses only, we will also exclude participants who fail manipulation checks. We also included a citation that supports the use of manipulation checks for increasing attentiveness without weakening treatment effects:

> Kane, J. V., & Barabas, J. (2019). No harm in checking: Using factual manipulation checks to assess attentiveness in experiments. American Journal of Political Science, 63(1), 234–249. https://doi.org/10.1111/ajps.12396

Participants answer a question before embarking on the survey that requires their understanding and consent to having attention/comprehension checks.

> **7. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses. The authors provided a clear design summary table (p.5), though some methodological aspects were less clear to me. Bastian et al. (2012) write in their results of Study 2 that "mean mental capacitity ratings were calculated for each animal and each condition. Participants' ratings of sheep and cows did not differ within either condition so we collapsed across versions. This yielded two animal types: food animal and nonfood animal." (p. 250).**
> **This implies that they originally did another test before they performed the t test. I suspect it was an ANOVA with a between-subjects factor (species: sheep last/first vs. cow first/last?) and a within-subjects factor (non-food vs. food), but this is not mentioned. Will the authors also test this between-subjects effect? How would the authors respond if participants' ratings of sheep and cows do differ? What would be the consequences for interpreting the results? In short, I think the experimental set-up of Study 2 requires more methodological detail. That being said, I laud the authors for their personal correspondence with the first author to confirm a description error concerning the t test.**

Thank you for your observation regarding the mixed analysis that the authors likely conducted. As an additional analysis, we added the suggested 2 X 2 mixed factorial ANOVA with animal food status (food vs. nonfood) as the within-subjects factor and animal species (cow-first vs. lamb-first) as the between-subjects factor and perceived animal mental capacities as the dependent variable. If the animal species order makes a meaningful difference, it would suggest that participants are judging cows and lambs differently and that perceptions of meat animals' minds vary by species and should be tested separately in future research. If the interaction is supported, it would suggest that effects might only arise in response to one animal but not another, and we are open to this possibility.

> **8. Concerning the attention checks, the exclusion criteria are rather vague; the authors merely note that failing the checks could be reasons for exclusion, so I recommend more clarity here.**

Thank you for pointing out this vagueness. We appreciate the feedback to improve on our exclusion criteria. To clarify, we will focus our analyses on the full sample. However, as a supplementary analysis and to examine any potential issues, we will also examine the results with exclusions if we fail to find support for the findings. In that case, we will report the exclusions in detail and will report results for both the full sample and with exclusions. Our criteria are now described in more detail in the manuscript and elaborated upon in the Supplementary Materials.

> **The authors combined the two studies into a singular data collection (displayed in random order and with minor adjustments) and mention that this design allows "to both test the designs of the original studies, and to run further tests in comparing the effects of the different studies with the potential of additional insights" (p. 10). I wonder which further tests the authors had in mind because I did read anything about this resolution in the method section.**

Thank you for pointing out this ambiguity. Please see our response to item #1 (the response to the editor).

> **9. As a sidenote, it might not be clear to participants whether "last week (p. 17)/month (p.26)" includes their participation in the questionnaire. Nevertheless, if the participants pay attention the chances seem high that they select "Used a computer, tablet, or mobile phone" anyway, so this does not seem to be a big issue. In addition, the first answer option (p.26) is in the present tense ("run a marathon") unlike the other options which are in the past tense.**

Thank you for this observation. This attention check has been used in published work (Jacobs & McConnell, 2022) without complaints or reported confusion from participants and is based on recommendations for constructing attention checks (Abbey & Meloy, 2017). Nonetheless, we are gladly revise given the feedback.

We changed the instructions to be clearer that the current session is included in “this week”. We also agree that the first answer option should be in past tense and we have made this change. We also have changed the instructions in the Supplemental Materials and Qualtrics programming to say “week” rather than “month”.

> **10. In the supplementary material, I think the sheep in the picture is incorrectly described as a lamb. I recommend to resolve this inconsistency.**

Thank you for the concern, but we believe this label is correct. We recognize that the authors in the original paper used the terms sheep and lamb interchangeably. We reached out to the authors to receive their materials, and this description and picture were taken verbatim from their materials. Thus, we wish to stay consistent with the original study. To respect your concern and to recognize the confusing nature of the labels, we added a few sentences on this issue to our new “Future directions and notes from reviews” section in our Supplementary Materials. We also modified the manuscript to consistently use the term “lamb” except when directly quoting the original studies.

> **To ensure high data quality, the authors plan to include two attention checks and manipulation checks (though see points 1C-3 and 1D-3 for criticism). The authors will also employ a Qualtrics fraud and spam prevention measures (e.g., reCAPTCHA, etc.) and several CloudResearch options in MTurk (i.e., Duplicate IP Block, Duplicate Geocode Block, Suspicious Geocode Block, Verify Worker Country Location, Enhanced Privacy, CloudResearch Approved Participants, Block Low Quality Participants, etc.), though it is not clear to me what these CloudResearch options mean.**

Thank you for the comments. These were indeed planned and are common practice when we conduct data collection on CloudResearch/MTurk. We now switched to Prolific to allow for targeting participants based on dietary preferences, and so the CloudResearch options are no longer relevant.

> **11. Bastian et al. (2012) write in a footnote that they also gathered a sample of vegetarians to examine whether vegetarians lack a motivated denial of food animal minds. Including such a control in Study 2 could strengthen the theory (i.e., assess whether motivated denial only occurs in meat-eaters), though the design of the study would admittedly become more complex.**

Thank you for the suggestion. We agree that this would be an insightful extension but that it goes beyond the current scope of the replication. We will add this idea to the “Future Directions” section of the manuscript after data collection, and have added a new section to the Supplementary Materials to catalog useful suggestions such as this.

Response to Reviewer #3: Prof. Florian Lange

DISCLOSURE: Florian Lange has served as peer reviewer for projects in the CORE team (on a different "Pichert and Katsikopoulos, 2008" replication). The corresponding author commented that he lists Florian Lange as a peer reviewer member in the CORE team lists.

> **12. The authors propose a Registered Report of two of the studies included in Bastian et al. (2012). They combine a correlational study and a within-subject experiment into one online survey. Apart from this combination, the proposed methods are very close to the original study and any deviations are clearly identified and motivated. Other strengths of the proposal include the very good rationale for conducting a replication, the compelling power analysis and sampling plan, and the transparent reporting of all materials.**
>
> **If I had to criticize something, it is that the information and materials are at times a bit scattered throughout the multiple parts of the manuscript and supplement. As a result, I have been confused about, for example, which exclusion criteria will be applied in confirmatory analyses. Please find my remaining comments below. I hope they are helpful in revising this manuscript and I am looking forward to seeing the results of this high-quality Registered Report.**

Thank you for your time, the positive notes, and the thoughtful feedback.

> **13. I would revise the very first sentence. Who is “we” here and what is meant by “care”? Without this being clear, the reader cannot judge if “we care for animals” is truly a “fact”? Perhaps it is rather that some people express appreciation of animals yet eat them. I also do not think that it would be reasonable to say that such an observation “demands explanation”.**

Thank you for the comment. We appreciate the note and agree that the usage of ‘we’ in this context is ambiguous and glosses over many important distinctions in moral beliefs about animals in those who eat meat and those who don’t. We are also happy to moderate the strong statement about ‘demanding’ an explanation. We attempted to reformulate in a way that is more precise about who the ‘we’ is in this context.

> “~~The fact that we care for animals yet eat them demands explanation.~~ **As a society, we care for animals yet eat them.** Loughnan et al. (2010) coined this phenomenon the ‘meat paradox’...”

We use ‘we’ in reference to society as a whole, where we think the claim is interpretable and more accurately captures the fact that the majority of people generally hold warm feelings towards animals yet also regularly eat meat.

> **14. The term “moral patients” is unclear and so is what it would mean for human-animal relationships to be “legitimate”.**

Thank you for the feedback. We amended accordingly in an attempt to more clearly convey what we mean:

> “…mechanism that is thought to play an important role in resolving this tension and maintaining the paradox is the motivated denial of food-animals' minds and therefore ~~status as moral patients~~ **their capacity to feel pain and be harmed** (Bastian & Loughnan, 2017; Loughnan & Davies, 2020). By positing that people are motivated to deny the minds of the animals they eat, Bastian et al. (2012) present ~~a powerful psychological explanation, and strong indictment, of human relationships with animals. Indeed, how can human relationships with animals be considered legitimate if perceptions of animals as moral patients solipsistically emerge from how people wish to use them?~~ **a psychological explanation of how we can care for animals and simultaneously eat them**. This idea has garnered substantial interest and has framed much of the psychological approach to meat consumption (Bastian & Loughnan, 2017; Dhont & Hodson, 2020; Loughnan & Davies, 2020; Rothgerber, 2014; Piazza, 2020). It therefore seems timely and worthwhile to revisit Bastian et al.’s (2012) seminal studies on the motivated denial of food-animals’ minds.”

> **15. Paradox framing: I can see how framing so-called attitude-behavior gaps like the present one as "paradoxes" can seem appealing, but actually, this does not seem very sound or reasonable to me. The described pattern would only be paradoxical if we conceived of beliefs/attitudes as absolute entities/categorical constructs: people either care for animals (or another attitude object) or not and if they care for them, all their behaviors would need to be directed at expressing this attitude in order not to be paradoxical. I don't think this is how we treat attitudes in psychological research. We consider them as dimensional (more or less pronounced) and we know that different attitudinal goals can conflict. If people say they "care for animals" and "animals should not be harmed", but then eat animals, well, I would say their positive attitude/caring for animals was not particularly strong or their behavior was (also) a function of another attitude. This does not seem very paradoxical to me. Against this background, I would like to encourage the authors to rethink if they want to reproduce the paradox framing here.**

Thank you for your thoughts. We appreciate where you are coming from and can see how the framing might fall down under careful conceptual scrutiny. However, addressing the potential shortcomings of the framing has to be weighed against the need to discuss the work within the context of the original author's views. We do not think this is the place for a careful critique of the framing of the meat paradox, and are concerned that by making the edits implied by the reviewers comment would take the manuscript in a direction that diverges from our primary purpose of replication. For this reason, we have opted not to make any edits in response to this comment.

> **16. I would not say that the Bastian studies "show that those animals that are perceived to be edible are also likely to be perceived as lacking a mind..." This would be an overgeneralization of their findings which have been obtained under very limited conditions – I don't think the phenomenon can be considered established yet.**

Thank you for the comment. We are happy to moderate and now describe such claims as arguments posited by the original authors (as opposed to empirically-established facts).

> *Bastian et al. (2012) presented two crucial tests of the perspective. In an initial study, they asked 71 students about their perceptions of 32 animals and found that the degree to which animals were perceived to be edible was positively related to beliefs that they lacked minds. In a follow-up study, they prompted 66 students to consider two animals, one which was destined to be taken to an abattoir and slaughtered for meat and one which was destined to be moved to a paddock and spend its time eating grass. The animal that was destined to be slaughtered for meat was perceived to possess a less sophisticated*

> *mind than the animal that was destined to be moved to a paddock.* ~~*These studies support the idea that how we perceive animal minds is directly related to their status as sources of food by showing that: 1) those animals that are perceived to be edible are also likely to be perceived as lacking a mind, and 2) making an animals' status as a source of food salient can lead people to perceive it as lacking a mind.*~~ ***On the basis of these findings, the authors argue that: 1) those animals that are perceived to be edible are also likely to be perceived as lacking a mind, and 2) making an animals' status as a source of food salient can lead people to perceive it as lacking a mind. Taken together, the studies suggest that how we perceive animal minds is directly related to their status as sources of food.***

> **17. I think the calculations and considerations regarding statistical power are very strong and compelling. I would only change the last sentence of that section as it is not clear to me for what "this sample size would be more than sufficient".**

Thank you for the praise regarding our power analyses. To avoid the unclear claim of sufficiency, we deleted the phrase you mentioned.

> **18. As to the payment part, it is not clear if the data from the 30 participants has led to an adjustment of the payment amount and, more importantly, if these data (which have already been collected, I think) will be included in the total dataset (so that only 970 additional participants will be recruited). Please clarify.**

The data from the 30 pretest will be included and has been accounted for when calculating the payment. As described in the manuscript, their data will be analyzed, and we only examine issues of pay, duration, and possible technical issues in the funneling section.

> **19. I do not understand the section on Manipulations and its function. This information has already provided elsewhere and if anything, it is more vague here than it has been before. In general, I find the structure a bit jumpy and the logic of the headings not always clear. One factor contributing to this might be that sometimes, the word "replication" is added to the heading (I don't clearly understand when and why).**

Thank you for pointing out this redundancy. As we agree that the information is better described elsewhere, we removed the "Manipulations" section. We also went through the headings and removed mentions of "replication" from headings, which were left over from a prior version of the manuscript that also included extensions.

> **20. I think the statements regarding the exclusion criteria contradict each other. Here, it says that participants who failed the attention checks will be excluded; on page 17, it says that such participants could be excluded in exploratory analyses; and I think the supplementary materials on page 28 contain the same contradiction. I would encourage listing all criteria that will be applied for the confirmatory analyses here (on page 21), not to refer to the supplement for this, and (potentially) not to refer to any criteria that might be applied in exploratory analyses to avoid confusion.**

Thank you for pointing this out. Please see our response to item #8 (reviewer 2).

> **21. Please make explicit: will 3 SD outliers (for which variables) be removed?**

Based on feedback received here and from other PCIRR peer reviews, we decided that the outlier analysis does not add a meaningful contribution to the replication and adds unnecessary decision-making “degrees-of-freedom”. Thus, we removed these outlier analyses.

> **22. I think it was not made explicit if results will be tested using one-sided or two-sided tests. Please specify and sorry if I overlooked it. I also did not find the evidence criteria. I assume the authors will consider the replication of Study 2 successful if $p < .05$ and the difference is in the same direction as in the original, but this should be made explicit. For Study 1, is this a family of hypotheses? What will be concluded if only one or two of the corresponding tests are significant? Will there be correction for multiple testing?**

Thank you for pointing out this lack of clarity. When conducting significance tests, we added that we are using two-sided tests. However, our evidence criteria are not based on statistical significance and are instead based on comparing effect sizes with the original article. As stated in the manuscript, our evidence criteria is specifically based on LeBel et al. (2018) and were detailed in the Supplementary Materials. For Study 1, we will consider it to be a successful replication if all three hypotheses are supported, a mixed replication if only one or two of the hypotheses are supported, and a failed replication if none of the hypotheses are supported. We do not plan to conduct multiple comparison adjustments because the hypotheses are fairly discrete and defined a priori.

> **23. Supplement: I think this needs to be cleaned up. Some of the parts do not seem relevant for and adjusted to the specific study (or I do not understand the relationship). The section "additional information about the study" seems to include methodological information that is missing in the main text. Please integrate.**

Thank you for this observation and for the opportunity to improve our supplemental materials. First we note the "additional information about the study" section is to be filled in after data collection. We have made this clearer in the text. Throughout the Supplemental Materials, any parts that are to be filled in later have been marked with a red label in a larger font reading "To be completed after data collection".

> **24. I appreciate the careful analysis of between-study differences. I think the removal of the filler task is worth another thought. I do follow the authors' reasoning (its theoretical significance has not been clarified in the original study, so the original authors probably did not consider it to be very important), but I guess it is not implausible for this methodological detail to play a critical role. The lack of a filler task might both promote consistency in participants' judgments or amplify the contrast between the two animals. I think it would be worthwhile to contact the original authors about this.**

Thank you for encouraging us to more carefully consider the filler task. Although we recognize that this is a change from the original study, we feel that the minimal description of the filler task makes it counterproductive to include a filler task. We do not know what the task was, and including a random filler task could create other inconsistencies between the studies that could add noise to the effects. Nonetheless, we have added a paragraph to our "Future directions and notes from reviews" section of the Supplementary Materials that reflects your concern and we will add it to the "Limitations" section of the general discussion after data collection.

> **25. In the participants section, I was missing a description of the sampling process and a priori exclusion criteria. From which population will the sample be a sample = to which population will/can the results be generalized? Will the study be advertised to everyone on MTurk and then filled on a first-come-first-serve basis? Will it be advertised specifically to non-vegetarians? I also stumbled across the "etc" in the MTurk/Qualtrics quality checks. Isn't it possibility to provide an exhaustive list at this point?**

Thank you for pointing out this lack of clarity. See our response to items #3 and #8. We will remove the "etc." and provide an exhaustive list of Prolific quality checks. Our sample will be non-vegetarians that are part of the Prolific subject pool and filled on a first-come-first-serve basis. This has been clarified in the manuscript.

Response to Reviewer #4: Prof. Sebastian Berger

My evaluation is that the scientific validity of the research question is high. The present submission is an attempt to directly replicate two studies from a paper by Bastian et al. (2012). The authors of the original study investigate the "meat paradox". The central idea is that meat takes a prominent place in most people's diet and is part of their culinary enjoyment. At the same time, people dislike harm done to animals, creating an inconsistency referred to as the meat-paradox (Loughnan et al., 2010). People's concern for animal welfare conflicts with their culinary behavior. The authors of the original study argue that people are motivated strongly to overcome this inconsistency and that this is achieved through motivated moral disengagement driven by a psychologically aversive tension between people's moral standards (caring for animals) and their behavior (eating them). In particular, they focus on one disengagement mechanism, namely the denial of food animal minds, and therefore their status as moral patients.

The original paper has been quite influential and has received 476 citations (google scholar, assessed May, 2022). Besides this obvious scientific impact, the underlying theoretical ideas of the paper have been influential with respect to a lot of follow-up research on meat consumption. I therefore think that it is a very suitable target for a replication and the authors of the present manuscript (henceforth: Jacobs et al. (2022)) attempt to subject Studies 1 and 2 of Bastian et al. (2012) to high-powered direct replications. I am reviewing the research project as a Stage 1 RR.

Thank you for the feedback and for the thoughtful comments.

1B. The logic, rationale, and plausibility of the proposed hypotheses, as applicable.

The present submission includes four hypotheses (1a-c, 2). I evaluate them jointly below, as my overall evaluation is very positive.

Hypothesis 1a: Mind attribution is negatively associated with perceived edibility of animals.

Hypothesis 1b: Mind attribution is positively associated with negative affect regarding eating animals.

Hypothesis 1c: Mind attribution is positively associated with moral concern for animals.

Evaluation: Hypotheses 1a-c are straightforward in my opinion. They directly follow from the original research that Jacobs et al. (2022) attempt to replicate. In Hypothesis 1a-c, there is no causal hypothesis, thereby the data will be mute with respect to whether eating an animal causes humans to deny a mind, or whether mind-attributions cause lack of eating. Given the existing research, the correlational hypothesis is sound.

Hypothesis 2: Being told that animals will be raised for food consumption (compared to being told it will live as a grazing animal) leads to denial of mind to those animals.

Evaluation: Hypothesis 2 follows from an experimental design able to assess the causal relationship (eating àmind denial). The hypothesis is straightforward and sound.

1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable).

The statistical power of the current research design is plausible. The authors have taken great care not to overestimate the originally reported effect size and go to great length justifying their sampling decisions, including necessary checks and balances. They adjust their power using the safeguard method and replace the original effect size with the lower bound of the 60% CI. In addition, the use the smallest effect size from Study 1 and 2 in the original study. Due to a change in the design (fewer trials), they further increase the required sample size to 1,000 respondents. Thus, power should be high enough to detect an effect, if any.

Further, as Jacobs et al. (2022) attempt to recruit participants online via Amazon Mechanical Turk, further steps to assure data quality are necessary. The authors use several measures to secure data integrity. The most relevant measure is the recruitment via Cloudresearch/Turkprime and several measures to employ high data quality.

I want to comment on other potential aspects. Note that many aspects raised result from the fact that the authors are attempting a close replication of Bastian et al. (2012), and potential weaknesses may arise from weaknesses that occurred in the original study, rather than in the replication.

26. Blinding: One issue may arise from experimenter demand effects. Asking about the suitability of various animals that are typically not part

> **of the diet in the target population (e.g., lion, elephant, etc.), it could be that participants easily guess potential research questions. Thus, a threat to internal validity is, in my opinion, that blinding participants to the research hypothesis is not easily possible. A potential solution could lie in asking participants at the end of the survey about the hypothesis or the research question. If such a question is asked, the exclusion protocol needs to be adjusted.**

Thank you for this observation. Although it is possible that participants will deduce that the study is researching perceptions of animal minds, we believe that is unlikely that they could determine the nuanced aspects of the research questions, namely that people deny minds to food animals in order to avoid dissonance, in a way that would affect their responses. We also note that past research suggests that there is little evidence that within-subjects designs are more vulnerable to demand effects in similar JDM studies (Aczel et al., 2018; Lambdin & Shaffer, 2009). Nonetheless, for any design, it is always important to account for potential blinding issues and demand effects, and therefore we include the following question in the funneling section at the end of the survey: “What do you think the purpose of the study was? (one sentence)”.

Aczel, B., Szollosi, A., & Bago, B. (2018). The effect of transparency on framing effects in within-subject designs. Journal of Behavioral Decision Making, 31(1), 25-39.https://onlinelibrary.wiley.com/doi/full/10.1002/bdm.2036

Lambdin, C., & Shaffer, V. A. (2009). Are within-subjects designs transparent?. Judgment and Decision Making, 4(7), 554-566.https://journal.sjdm.org/9921/jdm9921.pdf

27. Randomization: The randomization of Study 2 (i.e., the experimental study) seems unproblematic and order effects arising from the counterbalancing should be easily detectable, given the large sample size. However, I fear that a potential risk from unintended order effects may restrict the range of data in Study 1, potentially causing ceiling effects. I think I can most easily describe this risk from my own experience clicking through the survey. When taking the survey, I was first prompted to rate a mole. I ascribed particularly high values in the questions asking for mind-attribution, not having seen the other animals. Then, some animals were presented, which I am ascribing "higher" mind-attribution (e.g., elephant, gorilla, monkey): Hence, I would have produced more variance initially, had I known the range of animals. The same issue may arise with the likelihood of eating. In consequence, I think that the presentation order may strongly affect whether or not support for Hypotheses 1a-1c may be found.

Thank you for this observation. As stated on pg. 15, the order of the animals presented is randomized and which animals are presented is randomized. Any noise caused by the order of the animals should be spread out between participants and therefore should not strongly affect the strength of the effects. In addition, not randomizing the order of the could be problematic because it would only speak to effects for that particular order of presentation. Thus, we did not change the randomization of animals in Study 1 but will clarify that the order is indeed randomized. We do agree that it would be useful for participants to be aware of the full range of animals. Thus, in the survey, we have changed the instructions for Study 1 to let participants know the range of animals and that they are presented in no particular order.

28. On a more general level, I have a concern that beliefs in animal emotion and cognition could have been on the rise since the publication of the original study, thereby increasing the problem of ceiling effects. Many popular science books (e.g., Frans de Waal has published two books that I recall. "Are we smart enough to know how smart animals are?" and "Mamas last hug") have emerged on the topic. In addition, Netflix has devoted wide-received documentaries to the topic of animal intelligence and emotions, for example "My Octopus teacher" or "Explained: Animal Intelligence". In addition, animal welfare and vegetarianism is potentially more broadly discussed in the media than in 2010. As a result, a potential non-replication may result from general trends in the society, rather than from the absence of a true effect in in the initial study at the given time point.

Thank you for this observation; we agree that this is an interesting possibility. It's hard to say for sure, as we aren't aware of any longitudinal data tracking perceptions of animal minds. On a practical note, we are relatively confident that we are unlikely to encounter problems associated with ceiling effects, given the animals and measures we are proposing here. This is because recently conducted studies using similar animals and measures which drew on the same population as we are planning to (Prolific Academic) showed no evidence of ceiling effects (see for example Leach et al. 2021)

Rather than seeing this just as a challenge, we also see this as adding to the value of the replication by allowing us to compare changes over time. In response to your comment, we added a paragraph on this idea to the "Future directions" section of our Supplementary Materials, which will be added to the main manuscript after data collection. We also describe in the Supplementary Materials how we will also compare the descriptive statistics between the original studies and the replication to examine changes over time.

Leach, S., Sutton, R. M., Dhont, K., & Douglas, K. M. (2021). When is it wrong to eat animals? The relevance of different animal traits and behaviours. European Journal of Social Psychology, 51(1), 113-123.

> **29. A remaining risk factor is that the authors may be unable to recruit 1,000 participants within a given time-frame. Due to my inexperience with Cloudresearch/Turkprime, I cannot quantify this risk, but I want to suggest a risk mitigation plan that may include collection of additional data via an alternative research platform (e.g., Prolific), if necessary.**

Thank you for your concern. In our experience conducting these replication studies (15 total, 9 with an in-principal acceptance), there should be no trouble recruiting 1,000 participants on Prolific within a realistic timeframe. For example, past CORE studies on Prolific with 2,000 participants have finished collection in less than one day. Although we are using more specific criteria, Prolific lists over 19,000 available participants that match our needs.

> **30. The research question is a direct (or close) replication. Thus, the authors are constrained in freely deciding upon the experimental design. As discussed above, order effects may inflate the risk of ceiling effects. In general, I encourage the authors to include a mitigation plan for this possibility. Although this does make the stage 1 report a bit more complex, I think that it may help with respect to assessing the replicability of the original result. An alternative could be to run a more conceptual rather than direct replication, should the problem of ceiling effects emerge in piloting.**

Thank you for your thoughts here. Please see our response to item #1 (response to editor).

Deviations from research protocol with respect to original work: The authors are very transparent about the deviations in their research protocol. I evaluate all deviations from the original study as plausible and would support the updated study design. The authors do a good job making these very transparent.

1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses.

The clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures. Undisclosed flexibility in the procedure and analysis is unlikely to remain undetected.