

Peer Review and Communication History

MS Title: Analysis Blinding as a Potential Means to Foster a Productive Collaboration Between Original Authors and Replicators

Author Names: Alexandra Sarafoglou and Suzanne Hoogeveen

Submitted: Dec 11, 2024

Editor First Decision: Revise & Resubmit
Feb 10, 2025

Dear Alexandra Sarafoglou,

I have now received two reviews of your manuscript, “Analysis Blinding as a Means to Foster a Productive Collaboration Between Original Authors and Replicators”. I also independently read the manuscript before consulting these reviews. The reviewers and I had very positive reactions to your manuscript. There are only minor issues that I would like to address. Specifically, reviewer 1 pointed out that some of the language in the manuscript could be adjusted to clarify that this is an opinion piece and avoid the impression that the suggestion had been tested.

I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology. In your resubmission, please include a document with a point-by-point response to both the points I list here and the reviewers’ comments, outlining each change made in your manuscript or providing a suitable rebuttal. Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all necessary copyright permissions have been obtained. This may be the last opportunity for major editing, therefore please fully check your file prior to resubmission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,
Merle-Marie Pittelkow

Reviewer 1

Rating scale questions

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript)			✓		
The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript)			✓		
The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.)			✓		
The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.)			✓		

Open response questions

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

This perspective article presents how analysis blinding can be implemented, as a potential way to make discussions between original and replication authors most productive, especially after the replication analysis was conducted. The authors were involved in large replication effort, and explain the limitations of the collaborations they had with original authors. Their suggested approach, to blind the analysis by having an independent data manager alter the data before handing it to the analyst who will then conduct the analysis blinded. Analysis blinding is not new, and has been previously proposed to prevent questionable research practices in various fields. The application for collaboration between original and replication authors is novel and has lots of potential. I do have a couple of comments the authors could address, most of them regarding clarifications.

Major point: While the authors present how analysis blinding can be implemented, they do not actually establish a link between blinding and productive collaboration with authors, e.g., they did not test it. However, this is what their title and some of the conclusions claim as f.ex end of the abstract “We conclude that analysis blinding can prevent fruitless discussions and tension between original authors and replication teams..”. After reading the beginning of the manuscript, I thought the

authors ran a small trial to test the effect of blinding. I therefore would suggest to weaken the conclusions (and title) a bit, and make it obvious that those are speculations. Just to be clear, I share the authors hypothesis/believe, but this hypothesis remains to be tested.

Minor points:

- MTurk: Could the authors add some indication on what this is and why the team behind Holzmeister et al chose it. Especially because the next paragraph mentions data quality concerns related to the platform.
- It was not clear whether the discussions with the original authors were a mandatory part of the replication efforts. If this was not the case, in which cases or why would replication teams contact original authors.
- “Yet due to the nature of the replication project, we were unable to effectively address all concerns raised by the original authors” - are these concerns raised prior to data collection? and it is unclear what the “nature of the project” is that precluded addressing them.
- p2 line 48: a “to” is missing in “people tend evaluate evidence”
- p3 line 56: by whom were replication studies approved? and why were original authors then still contacted even if the protocols could not be altered anymore?
- p 4 para starting in line 97: you often refer to “authors” and mean “original authors”. I would add original everywhere, to make it absolutely clear who you are talking about.

Rachel Heyard (Postdoctoral fellow, Center for Reproducible Science, UZH)

Reviewer 2

Rating scale questions

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript)			✓		
The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript)			✓		
The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.)			✓		

Strongly
Disagree

Disagree

Neutral

Agree

Strongly
Agree

The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.)



Open response questions

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

I think this paper is great! It offers a simple, clear suggestion, one which I wholeheartedly endorse. It's clearly and compellingly written, with attention to nuance. (I was one of the authors whose work was being replicated, and I appreciate the replication team's consistent care in the process!) I recommend acceptance.

My only thought (which the authors can address if they want) is that, in this particular case, neither analysis blinding nor pre-registering the design would have addressed what was, to me, the biggest disagreement in the replication process: whether it made sense to use MTurk data filtered only by comprehension checks + IP address checks, given the precipitous decline in MTurk data quality between the time of the original studies & replication attempt. This was something that the authors & I disagreed about before the design was pre-registered, and is not something that would arise in the post-hoc analysis process (regardless of blinding or not). Of course, analysis blinding is not going to solve all the difficulties of the replication process, and the authors don't claim that it will. But given that the authors specifically mention this as one of the points of tension between them and the original authors, it seemed worth noting that this particular issue would not be addressed by the proposed solution.

Like I said, I think this paper is great, and think it is acceptable as is, regardless of whether the authors address this point. And again, I appreciate all the effort the authors put into the replication project!

Sincerely,
Adam Morris

Author Response

Feb 18, 2025

Dear Dr. Pittelkow,

My co-authors and I are happy to submit the revision of our manuscript “Analysis Blinding as a *Potential* Means to Foster a Productive Collaboration Between Original Authors and Replicators” for publication in *Collabra*.

We would like to thank the reviewers for the positive comments and the constructive feedback. We now made sure that it is clear this piece is a perspective, not empirical research, for instance, by changing the title and revising some statements across the article. We have also added more details on how the replication project was organized and the extent of the original authors’ involvement. Additionally, we acknowledge that analysis blinding will not resolve all tensions between replicators and original authors, particularly those arising prior to data collection. We also discuss how we believe future replication attempts can address this challenge. Below we have provided a detailed point-by-point response (in bold) to the comments by the reviewers. The adjusted passages in the manuscript are indicated in blue and copied to this response letter.

We are happy to incorporate the reviewers’ suggestions and we are confident that these changes considerably improved the quality of the manuscript. We look forward to your decision.

Comments by Reviewer 1

[R 1.1] This perspective article presents how analysis blinding can be implemented, as a potential way to make discussions between original and replication authors most productive, especially after the replication analysis was conducted. The authors were involved in large replication effort, and explain the limitations of the collaborations they had with original authors. Their suggested approach, to blind the analysis by having an independent data manager alter the data before handing it to the analyst who will then conduct the analysis blinded. Analysis blinding is not new, and has been previously proposed to prevent questionable research practices in various fields. The application for collaboration between original and replication authors is novel and has lots of potential. I do have a couple of comments the authors could address, most of them regarding clarifications.

Major point: While the authors present how analysis blinding can be implemented, they do not actually establish a link between blinding and productive collaboration with authors, e.g., they did not test it. However, this is what their title and some of the conclusions claim as f.ex end of the abstract “We conclude that analysis blinding can prevent fruitless discussions and tension between original authors and replication teams..”. After reading the beginning of the manuscript, I thought the authors ran a small trial to test the effect of blinding. I therefore would suggest to weaken the conclusions (and title) a bit, and make it obvious that those are speculations. Just to be clear, I share the authors hypothesis/believe, but this hypothesis remains to be tested.

Thank you for pointing this out. We agree with the reviewer that readers should not be misled into thinking this is empirical research rather than an opinion piece. To clarify that we are presenting a perspective in this manuscript, we have made several revisions.

First, we have revised the title to: “Analysis Blinding as a *Potential* Means to Foster a Productive Collaboration Between Original Authors and Replicators”. Additionally, we have updated the abstract to reflect this distinction, as follows:

[...] We discuss three replication studies we were responsible for in the Holzmeister et al. (2024) project and demonstrate how to effectively blind data for each of them. We *argue* that analysis blinding *has the potential to* prevent fruitless discussions and tension between original authors and replication teams in replication projects while preserving a healthy scientific debate.

We have also weakened the following statement in the conclusion section on page 6:

We believe any additional time and effort related to blinding the data is more than compensated for by the time and effort *it may save* on unproductive discussions between the original authors and the

replicators. It gives the original authors a fair chance to voice concerns without the risk of being perceived as trying to manipulate the results in their favor. Above all, in our view, analysis blinding fosters a more collaborative environment and hence *may* lead to a more satisfying experience for all parties involved in replication projects.

Finally, we included a reference of empirical evidence comparing preregistration and analysis blinding on page 4:

The resulting analysis is fair because the analysis pipeline is shielded against potential biases driven by the eventual results. At the same time, the resulting analysis is also flexible in the sense that the analytic strategy can be adapted to peculiarities even after data collection. This flexibility offers a key benefit as Sarafoglou et al. (2023) found: analytic strategies using analysis blinding lead to fewer deviations than those relying on preregistered plans. In the context of replication projects we propose to use preregistration and analysis blinding in combination. Specifically, a feedback round for the original authors could be implemented after data collection but before revealing the final data.

[R 1.2] MTurk: Could the authors add some indication on what this is and why the team behind Holzmeister et al chose it. Especially because the next paragraph mentions data quality concerns related to the platform.

We have now added a brief clarification that MTurk refers to an online data collection platform, along with an explanation of why Holzmeister et al. chose this platform for their replication attempts, page 2:

The aim of the Holzmeister et al. (2024) project was to replicate social science studies that were published in the Proceedings of the National Academy of Sciences (PNAS) between 2015-2018 and that were conducted on the online data collection platform Amazon Mechanical Turk (MTurk). The replications were high-powered, preregistered, and kept as similar as possible to the original studies. The team leaders' strict emphasis on direct replication ensured that the sampling platform, instructions, materials, and analysis remained identical to those of the original study. This was done to protect the project from concerns that hidden moderators might influence the results of the replications. Deviations from the original protocol were not allowed, with changes implemented only in exceptional cases (e.g., if an error was discovered in the original analysis).

[R 1.3] It was not clear whether the discussions with the original authors were a mandatory part of the replication efforts. If this was not the case, in which cases or why would replication teams contact original authors.

We have added clarification regarding the involvement of the original authors on page 2. Specifically, the original authors were invited to provide feedback and approve the replication protocols, and this process was followed for all replication attempts.

The original authors were involved in providing feedback and approving the replication protocols. These discussions took place before data collection for all replication attempts, with original authors explicitly asked to confirm the correspondence of the replication protocol to the original study. When original authors raised concerns, their arguments were considered before the lead team made a final decision. If the original authors did not respond after multiple follow-up emails, the lead team allowed the replication teams to proceed without permission. After the replication was completed and the data were analyzed, the original authors were contacted again to review and approve the final replication report summarizing the results.

[R 1.4] “Yet due to the nature of the replication project, we were unable to effectively address all concerns raised by the original authors” - are these concerns raised prior to data collection? and it is unclear what the “nature of the project” is that precluded addressing them.

We have now provided more details on how the discussions and involvement with the original authors took place (see also [R 1.3]). In response to the question of whether concerns were raised before or after data collection, and how these concerns were addressed, we explain on page 2–3:

The original authors were involved in providing feedback and approving the replication protocols. These

discussions took place before data collection for all replication attempts, with original authors explicitly asked to confirm the correspondence of the replication protocol to the original study. When original authors raised concerns, their arguments were considered before the lead team made a final decision. [...] After the replication was completed and the data were analyzed, the original authors were contacted again to review and approve the final replication report summarizing the results. [...] Yet for any issues raised after data collection, it was impossible to consider the original authors' arguments in an objective manner, as both parties were already aware of the replication's success or failure.

We also revised the statement on page 3 to make it less ambiguous:

On the other hand, many issues were raised by the original authors when they reviewed the final replication report – that is, after they had seen the replication results. This timing, in combination with the project setup of conducting direct, pre-approved replications, severely limited our flexibility in incorporating original authors' perspectives, however reasonable given the situation at hand.

Additionally, we have revised the section on “Potential Risks When Disputing Replication Results”, where we now discuss more concretely to what extent we believe replicators should remain sensitive to reasonable adjustments suggested by the original authors, page 4:

The core problem is not differing opinions on what constitutes a valid replication (e.g., whether changes in the subject pool or the “state of the world” affect comparability), but rather the challenge of addressing

data-dependent concerns in an unbiased way. We believe that even in the context of strict direct replications, some flexibility in protocol deviations should be maintained—if justified independently of the key statistical outcomes. For instance, if comprehension rates show a significant decline from the original study, a decision how to proceed (e.g., revise inclusion criteria) should be considered without knowledge of the replication results.

*[R 1.5] p2 line 48: a “to” is missing in “people tend evaluate evidence”
Corrected.*

[R 1.6] p3 line 56: by whom were replication studies approved? And why were original authors then still contacted even if the protocols could not be altered anymore?

See answer to [R 1.3].

[R 1.7] p 4 para starting in line 97: you often refer to “authors” and mean “original authors”. I would add original everywhere, to make it absolutely clear who you are talking about.

Thank you for pointing this out. We agree with the suggestion and now consistently refer to “original authors” throughout the perspective.

Comments by Reviewer 2

Comments to the Author

[R 2.1] I think this paper is great! It offers a simple, clear suggestion, one which I wholeheartedly endorse. It's clearly and compellingly written, with attention to nuance. (I was one of the authors whose work was being replicated, and I appreciate the replication team's consistent care in the process!) I recommend acceptance. My only thought (which the authors can address if they want) is that, in this particular case, neither analysis blinding nor pre-registering the design would have addressed what was, to me, the biggest disagreement in the replication process: whether it made sense to use MTurk data filtered only by comprehension checks + IP address checks, given the precipitous decline in MTurk data quality between the time of the original studies & replication attempt. This was something that the authors & I disagreed about before the design was pre-registered, and is not something that would arise in the post-hoc analysis process (regardless of blinding or not). Of course, analysis blinding is not going to solve all the difficulties of the replication process, and the authors don't claim that it will. But given that the authors specifically mention this as one of the points of tension between them and the original authors, it seemed worth noting that this particular issue would not be addressed by the proposed

solution. Like I said, I think this paper is great, and think it is acceptable as is, regardless of whether

the authors address this point. And again, I appreciate all the effort the authors put into the replication project!

We thank the reviewer for the positive comments and agree with their suggestions. In the conclusion section, we now acknowledge that while analysis blinding can help mitigate some conflicts between replicators and original authors, it cannot resolve all potential disagreements. However, this methodology can promote a fair approach to data analysis that aligns with the interests of both replicators and original authors. For example, in cases where data quality was a concern, analysis blinding could have allowed original authors to objectively assess the data and determine whether the replication attempt using the MTurk sample was indeed insufficient. We now write the following on page 6:

Analysis blinding is not a panacea – it may not, for instance, resolve fundamental differences in perspective between replicators and original authors on what qualifies as a good (direct) replication. In the Holzmeister et al. (2024) project, the lead team considered the data collection platform (i.e., MTurk) an essential part of the replication. However, some original authors argued that changes in the subject pool on MTurk rendered the newly collected data incomparable to the original data. Such divergent opinions should ideally be resolved beforehand, or else accepted as inherent disagreements. Nevertheless, analysis blinding could still add an opportunity to assess the data quality and subsequently make adjustments that both replicators and original authors agree upon.

Additionally, we revised the “Potential Risks When Disputing Replication Results” section (see also [R 1.4]) to discuss the balance replicators must strike between maintaining the direct nature of the replication attempt and remaining sensitive to reasonable adjustments, page 4:

The core problem is not differing opinions on what constitutes a valid replication (e.g., whether changes in the subject pool or the “state of the world” affect comparability), but rather the challenge of addressing data-dependent concerns in an unbiased way. We believe that even in the context of strict direct replications, some flexibility in protocol deviations should be maintained—if justified independently of the key statistical outcomes. For instance, if comprehension rates show a significant decline from the original study, a decision how to proceed (e.g., revise inclusion criteria) should be considered without knowledge of the replication results.

References

*Holzmeister, F., Johannesson, M., Camerer, C. F., Chen, Y., Ho, T.-H., Hoogeveen, S., Huber, J., Imai, T., Jin, L., Kirchler, M., Ly, A., Mandl, B., Manfredi, D., Nave, G., Nosek, B. A., Pfeiffer, T., Sarafoglou, A., Schwaiger, R., Wagenmakers, E., . . . Dreber, A. (2024). Examining the replicability of online experiments selected by a decision market. *Nature Human Behaviour*, 1–15. <https://doi.org/10.1038/s41562-024-02062-9>*

*Sarafoglou, A., Hoogeveen, S., & Wagenmakers, E.-J. (2023). Comparing analysis blinding with preregistration in the Many-analysts Religion Project. *Advances in Methods and Practices in Psychological Science*, 6 (1), 25152459221128319. <https://doi.org/10.1177/25152459221128319>*

Editor Final Decision: Accept

March 28, 2025

Dear Alexandra Sarafoglou,

I have now had a chance to read over your manuscript "Analysis Blinding as a Potential Means to Foster a Productive Collaboration Between Original Authors and

Replicators”, along with the letter describing the changes you made. Thank you for your responsiveness to the concerns that the reviewers and I raised. I also asked the reviewer whose comments informed most of your changes to evaluate your responses.

I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work, I think it will make an important contribution to the literature and I look forward to seeing it published! I hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point.

You will be receiving separate correspondence regarding any production and technical comments, data deposits, as well as publication charges. We work with the Copyright Clearance Center to process any applicable APC charges. Please note that your APC transaction must be completed before your article gets published.

You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely,
Merle-Marie Pittelkow

Reviewer 1

Rating scale questions

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript)			✓		
The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript)			✓		
The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.)			✓		
The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts)				✓	

Strongly
Disagree

Disagree

Neutral

Agree

Strongly
Agree

(including from lab to real world), other populations, other stimuli or measures, etc.)

Open response questions

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

I thank the authors for addressing all my comments. and have no further comments.