Dear Dr. Conry-Murray, Dr. Feldman and Dr. Majumder,

thank you very much for your positive evaluation and helpful feedback on our registered report “The influence of offender identifiability in Second- and Third-Party Punishment” and the invitation to submit a revision. We are very grateful for your comments and suggestions and have summarized how we have dealt with each of the specific points raised in the following.

We did our best to comply with all requests for revision and believe the manuscript and planned experiment have been improved by incorporating your comments and feedback.

Note: original comments precede our answers. Page numbers listed in our response refer to locations in the revised manuscript, without the "tracked changes function" (uploaded to the OSF platform: https://osf.io/h5uys/?view_only=60348d468d8949aab31e9304128ec76a).

RECOMMENDER (Clare Conry-Murray):

1. Both reviewers expressed concern with the way the victim identifiability findings were summarized on p. 5. I suggest you consider both reviewers comments and then revise this section to address the publication bias in the metanalysis, which may help explain why victim identifiability did not replicate. I also suggest that you address what this means for the importance of the topic.

   Response:
   As detailed in our responses to comments #7 and #17, we have thoroughly revised the relevant paragraph to better reflect the current state of research on this topic (see pp. 5-6).
   The revised paragraph now reads as follows:

   > “Research on the identifiability effect originated in 1968 when Schelling first described the phenomenon that the death of a single individual life triggers stronger emotional reactions and helping behavior than the death of an anonymous, so-called "statistical" life. In essence, Schelling summarized his idea as "the more we know, the more we care" (Schelling, 1968). In line with this reasoning, a large body of research has tested the hypothesis that the mere identification of a victim by “irrelevant information” such as name, age, or a picture leads to stronger emotional reactions (e.g., empathy) and willingness to help (e.g., donation behavior) compared to cases where the victim remains anonymous (e.g., Hou et al., 2023; Kogut & Ritov, 2005a; Lee & Feeley, 2018; Small et al., 2007) . One theoretical explanation that has been proposed for this effect is that, by making the individual more concrete and increasing their psychological closeness (Mentovich et al., 2016; Pałka et al., 2023), identifiability amplifies positive and benevolent emotions, and, ultimately, increases positive behavioral intentions concerning the victim, such as helping behavior (Kogut et al., 2018; Kogut & Ritov, 2005a; Oppenheimer & Olivola, 2011). Of note, recent studies have raised questions about the robustness and strength of the victim identifiability effect. Maier and colleagues (2024), for example, conducted a re-analysis of meta-analytic evidence on the identifiable

> victim effect and reported moderate evidence of publication bias, and strong evidence for the absence of the average identified victim effect. Additionally, several recent studies have failed to replicate the victim identifiability effect (e.g., Majumder et al., 2022; Moche & Västfjäll, 2021; Vu et al., 2024; Wiss et al., 2015), further complicating the understanding of its strength and the conditions under which it operated in the literature. Given this mixed evidence, the extent and boundary conditions of the victim identifiability effect remain uncertain."

Importantly, as also outlined in our responses to comments #7 and #17, this mixed evidence does not reduce but rather increase the importance of the planned (pre-registered) study, as this literature certainly benefits from pre-registered research.

2. Like Dr Feldman, I wondered about how you chose the features of identifiability. Why might including a photo, name and age matter, out of all the ways an offender could be identified? I see that you will control for gender. Will you also control for race or any other features?

   Response:
   We chose to provide the offender's name, age, and picture as identifying information because these information are often reported in media coverage on offenses and provided in research on identifiability effects (see e.g., Genevsky et al., 2013; Kogut, 2011a; Morvinski, 2023, Wiss et al., 2015). We have added a footnote in the revised manuscript explaining this choice (footnote 9, p. 18):

   > "We provide name, age, and a picture as these elements are often reported in media coverage on offenses and provided in research on the identifiability effect (see e.g., Genevsky et al., 2013; Kogut, 2011a; Wiss et al., 2015)."

   Regarding the control for race and other features: Based on your comment, we decided to use a different data base for our pictures in which different ethnicities are represented: the Chicago Face Database (Ma et al., 2015). To address potential race effects, we have randomly selected three male and three female photos from each of the following ethnic groups: Asian, Black, Latino, and White. This selection includes individuals aged between 24 and 29. Consequently, our sample of pictures consists of 12 male and 12 female pictures and we will randomly present one of these to participants. Additionally, we have opted for race-neutral first names (see materials document & manuscript p. 18). We believe these measures ensure that our study remains focused on the concept of identifiability, without being confounded by features such as perceived or assumed race.

3. I thought that H1 could have been written more clearly. By "weaker intention to punish" do you mean a smaller effect or a reversed effect where identified would be punished less than unidentified in the 3rd party condition?

   Response:
   Thank you for highlighting this. We have revised the exact wording of Hypotheses

H1-H4 to provide greater clarity (pp. 13-14). The updated hypotheses now read as follows:

> "**H1:** The influence of offender identifiability on punishment intentions depends on the role of the punisher in the initial offense. Victims indicate stronger intentions to punish an identified compared to an unidentified offender. This effect should be reversed for third parties, who are expected to indicate weaker intentions to punish an identified compared to an unidentified offender.
> **H2:** The influence of offender identifiability on empathy toward the offender depends on the role of the evaluating individual in the initial offense. Victims express less empathy toward an identified compared to an unidentified offender. This effect should be reversed for third parties, who are expected to express more empathy toward an identified compared to an unidentified offender.
> **H3:** The influence of offender identifiability on moral outrage toward the offender depends on the role of the evaluating individual in the initial offense. Victims express more moral outrage toward an identified compared to an unidentified offender. This effect should be reversed for third parties, who are expected to express less moral outrage toward an identified compared to an unidentified offender.
> **H4:** The influence of offender identifiability on offender blaming depends on the role of the evaluating individual in the initial offense. Victims blame an identified offender more compared to an unidentified offender. This effect should be reversed for third parties, who are expected to blame an identified offender less compared to an unidentified offender."

4. Both reviewers mentioned that you might want to revise the way you treat the mediation analysis. I agree with them (especially concerns about power for the mediation analysis), but I also see the mediation analysis as a valuable way to investigate how these factors might contribute differently. Your hypotheses are all very similar, in that victims will judge the offender more negatively than third parties. If these DV are each uniquely useful to building a theory, it would be helpful to show how that they go beyond more or less negative attitudes. Your 5th hypothesis might shed light on how these factors differ if you predicted or explored more specific effects. My suggestion is to use the mediation analysis as an exploratory test.

   Response:
   We appreciate this insightful feedback. We agree that the mediation analyses comes with some problems while also offering some merits. Based on your suggestion, we have decided to remove the mediation hypothesis and analysis from our revised Stage 1 Registered Report but plan to conduct mediation analyses exploratively. Following PCI's recommendation to avoid including exploratory analyses in the Stage 1 report, we have completely removed both the hypothesis and the mediation analysis from the revised manuscript.

5. There is only one trial to test the effect, despite that you note that context is likely to have an effect on judgments. I wonder why you chose pickpocketing and what

features of pickpocketing might differ from other crimes in ways that could impact punishment and other relate judgments. I suggest you add more justification for selecting that particular crime, and later consider discussing how different crime might differ (which could be in the future-discussion).

Response:
We chose pickpocketing as it typically does not involve direct confrontation between the victim and the offender in the situation of the offense (as compared to, for example, an assault). This allows for the offender to be presented as either identified or unidentified in both the victim and third-party conditions. We have clarified this in the manuscript and plan to discuss how different crimes might elicit varied responses in the discussion section for future research (p. 17). The added explanation in the manuscript reads as follows:

> "We chose pickpocketing because this typically does not involve direct confrontation between the victim and the offender in the situation of the offense (as compared to, for example, an assault). Therefore, this scenario allows the offender to be presented as either identified or unidentified for both the victim and third-party role."

6. Check the language of a "hard treatment." I am a native English speaker and I agree with Dr Feldman that a "hard treatment" does not sound right. I do think you can say wither "angry at" or "angry with" or at least both sound correct to me.

   Response:
   Thank you very much! As also outlined below in our response to Comment #15c, we adjusted this in the materials and manuscript (pp. 18-20). We further took this opportunity to thoroughly review the entire materials again to ensure there are no other language issues.

Reviewer # 1 (Gilad Feldman):

7. Requests:
   1. Be more humble and cautious about this effect.
   2. Cite not just the classics, but also the more recent mixed/failed evidence.
   3. Be very clear about focusing on identifiability, and disentangle that from citations that conflate singularity or scope insensitivity.

   I think you might have similar issues with offender identifiability. I would generally suggest caution and humility looking at the evidence from the classics and aim to focus on more recent well-powered pre-registered studies, and leave open the question of what to expect here .

   Response:
   Thank you for this valuable feedback. In response, we have revised the relevant paragraph on page 5 to incorporate recent studies, including those summarized in Majumder et al. (2022). This revised paragraph now reflects a more nuanced view of

the victim identifiability effect, emphasizing its context-dependent and ambiguous nature. We have quoted the revised paragraph in our response to comment #1.

Thank you also for your helpful suggestion to disentangle identifiability from singularity or scope insensitivity! We have now clearly distinguished the identifiability effect (i.e., identification via "irrelevant information") from the singularity effect in the manuscript and emphasize that we are investigating identifiability, while keeping the offender singular (footnote 1, p. 5). Moreover, we double checked the literature we review in our manuscript to ensure that we refer to research on the identifiability effect only (but not studies focusing on singularity or scope insensitivity).

Lastly, with regard to the literature on the effect of offender identifiability: This literature is comparatively scarce. However, we carefully reviewed this literature with regard to the evidence reported here. While not all studies cited in our manuscript on this effect are preregistered and highly powered, we have no reason to expect any systematic biases in this literature on offender identifiability, especially as the evidence reported in this literature is already rather mixed. Nevertheless, we agree that it is particularly important to conduct preregistered research to add to this literature. Still, we will certainly keep this in mind when discussing our findings in light of the results reported in the literature.

8. Identifiability:
    a. Looking at your materials , like in the mixed evidence literature, your identifiability is not just identified vs. unidentified, but also whether there is a photo. This is likely to create a strong effect and impact, do note that you will not be able to disentangle the use of photo from identifiability. No need to add another condition, this is already ambitious, just need to note this and be cautious about how to interpret the differences between the two conditions and what it is that they manipulate.
    b. Beyond that, I don't have much to say about the introduction and/or value, other than that it's interesting.

Response:

a. Thank you for this comment! As detailed in our response to comment #2, we included a picture as identifying information for applied and theoretical reasons. From a theoretical perspective, adding a picture is yet another identifying information that has often been used in the literature (as name and age; see e.g., Genevsky et al., 2013; Kogut, 2011a; Morvinski, 2023, Wiss et al., 2015), although we agree that this is a quite strong manipulation of identifiability. Additionally, we think that it is important to draw randomly from several pictures that have been validated (Ma et al., 2015). We now explain the rationale behind adding a picture in our manuscript (see footnote 9, p. 18) and will consider this aspect in the interpretation and discussion of our results, as suggested.

b. Thank you!

9. Model
    a. There is a lot going on this in model, with moderation and mediation (several suggested mediators).
    b. This needs a lot of power, I don't think you're powered for that, more on that below.
    c. A diagram summarizing all the links in one figure would be helpful to summarize the already good layout of the hypotheses.

Response:

a. As per our response to comment #4 (but also see our responses to comments #18 and #19), we have decided to remove the mediation hypothesis and analysis from the Stage 1 protocol. However, we still plan to conduct these possible mediation analyses exploratively. Importantly, we follow PCI's recommendation to avoid including exploratory analyses in the Stage 1 report by removing the hypothesis and planned analyses from the revised manuscript.
b. Thus, the concern about insufficient power for the mediation analysis is now addressed.
c. We have included a figure displaying our interaction hypotheses to provide a clear overview (see Figure 1, p. 14).

10. Power analysis
    a. Your power analysis doesn't take into consideration the mediation model. I know it's complex, but if you can't tackle that well, might want to acknowledge that, and focus the power analysis on the core hypotheses, and make the mediations secondary and exploratory. I have doubts about this sample being able to tackle such a complex model.
    b. Reproducibility: Please include the script/screenshots from the power analysis.
    c. Gpower sometimes has issues with interactions (especially mixed, which is not the case here), and so can sometimes be underpowered. You might want to double-check with R packages/tools like ANOVA/super-power: https://shiny.ieis.tue.nl/anova_power/ that would run the simulation for you. https://github.com/arcaldwell49/Superpower ; https://aaroncaldwell.us/SuperpowerBook/ and if you pursue that please make that reproducible code/screenshots so that it's clear what was the input and output of the analysis.

Response:

a. As outlined above, we have removed the mediation hypothesis and analysis from our analysis plan. As recommended and detailed in the manuscript (see p. 15), we have focused our power analysis on our primary hypothesis (H1).
b. Based on your suggestion in comment #10c, we have decided to perform our a priori power analysis using R exclusively and removed the power analysis previously conducted with GPower from our manuscript (see our response to comment #10c below).

c. Thank you for pointing us to these tools. We have now performed the a priori power analyses in R using the "pwr" package and report this analysis and its results in our revised manuscript. This analysis is more conservative than GPower, as it does not consider the numerator df. Consequently, this analysis leads us to plan for a slightly larger sample size, as detailed in our revised manuscript (pp. 15-16). The revised a priori power analysis section now reads as follows:

> "We conducted an a priori power analysis to determine the required sample size to test our central hypothesis (Hypothesis 1) using the R package "pwr" (Champely, 2020). To ensure sufficient power for our study, we based our analysis on the effects of offender identifiability on punishment reported in the literature, ranging from small (Barak-Corren & Lewinsohn-Zamir, 2019, $f = 0.12$; Lewinsohn-Zamir et al., 2017, $f = 0.10$), to medium-sized effects (Small & Loewenstein, 2005, $f = 0.26$; Kogut, 2011b, Study 2, $f = 0.32$). However, given that this prior research only focused on either the victims or third parties as punishers, this only corresponds to the planned follow-up analyses (see below), but not our planned ANOVA testing the interaction effect of offender identifiability and role of the punisher on punishment intentions. Consequently, we planned to collect enough data to detect a small effect ($f = 0.125$) in a 2 x 2 between-subjects-design ANOVA, with a power of $1-\beta = .90$ and $\alpha = .05$. An a priori power analysis resulted in a required sample size of $N = 911$ participants. To account for dropouts or data exclusions based on inattentive participation (see below for more information on data exclusions), we aimed to recruit $N = 1000$ participants finishing the study up until the last relevant item (so-called "use-me" item, see below). The code to replicate this a priori power analysis can be found in our analysis script on the OSF."

11. Exclusion:
Please state clearly which participants you'll exclude. Is it index >2 for too fast?

Response:
An index > 2 indicates that participants responded too quickly. In detail, this index is calculated based on median completion times across the survey. An index > 2 means that, across all pages of the survey, a participant completed the pages at least twice as fast as the typical participant of this survey. We have adjusted this in our revised manuscript (footnote 6, p. 16):

> "The online platform SosciSurvey provides an index of relative completion speed. This index is calculated based on median completion times across the survey. An index > 2 means that, across all pages of the survey, a participant completed the pages at least twice as fast as the typical participant of this survey (Leiner, 2019)."

12. Attention checks:

a. From what I see in the materials, these appear on Page 8. If this doesn't come with the scenario, then this might be more of a memory test than an attention test.
b. Please indicate the attention checks clearly in the manuscript. Looked like you're expecting 10% exclusions, is that your common experience with these checks?

Response:

a. We agree that answering these questions correctly needs both sustained attention as well as some memory capacity. After careful discussion, we decided not to place the attention checks immediately after the scenario to avoid inducing demand effects. Given the short duration of the study, we are confident that (attentive) participants will be able to answer these questions correctly.
b. We have added more information on the attention check items in our revised manuscript (p. 20). We plan to collect an additional sample size of 10% based on our experiences from previous studies and dropout and exclusion estimates reported in the literature (Hoerger, 2010; Meade & Craig, 2012). We have added this information to the revised manuscript (footnote 5, p. 16).

13. Materials:
    a. Any reason not to display the questions together with the situation on the same page? Participants have a tendency to move forward quickly, would they have the chance to go back and recall, if not – why not show these on the same page?
    b. You wrote "In the offender identified condition, offenders will be identified by name, age, and picture (taken from the Basel Face Database; Walker et al., 2018). Names, ages, and pictures will be randomly selected from six possible variations each. Gender of the offender will be counterbalanced. ". In the supplementary I see only one, maybe I missed something. Please, make =everything= you're planning to run available. We can't help you ensure this is the best plan unless we see everything you're planning.
    c. English: It looks like you mostly used items from published scales, but some of it reads a bit off, so perhaps it's my bad English or it's lost in translation. I'm not a native speaker. I would recommend carefully examining some of the questions or simply using similar questions used and validated in this literature in English. For example: "hard treatment", maybe "harsh treatment"?. "I am angry at the offender" I think might be better with "I am angry with the offender", as some websites write "Angry with is used when referring to people or animals… Angry at is used when referring to objects or nouns (not people or animals)", but it might not matter, not sure. In any case, just suggesting that you double-check.

Response:

a. Thank you for this suggestion. We agree with this adjustment and have revised the materials accordingly. Both the situation and the questions will be displayed on the same page for the participants (see materials document

on OSF: https://osf.io/h5uys/?view_only=60348d468d8949aab31e9304128ec76a

b. Thank you for this comment. Initially, we uploaded one example only due to copyright issues associated with the Basel Face Database. As detailed in our response to comment #1, we have decided to use pictures from another validated database: The Chicago Face Database (Ma et al., 2015). Unfortunately, we still do not have permission to share these photos publicly. To address this, we have added the exact identifiers of the photos we are using to the Materials document. Interested readers can request access to the pictures directly from the database, a process which is straightforward and quick for scientific use. Moreover, you can view our pretest version of the questionnaire (https://survey.ifkw.lmu.de/JudgmentRevealed/?act=1Zng3r85AfE2K8wgGWd1wSEc) to get an idea of the materials we are using. Although this approach is not ideal, it helps us comply with copyright regulations. Despite these limitations, we are confident that the validated images from the Chicago Face Database are precisely suited for the needs of this study.

c. We have thoroughly reviewed the items and instructions in our study. Based on your suggestions, we have made the necessary adjustments (see materials document and manuscript pp. 18-20).

14. Planned analyses

a. I would strongly recommend that you simulate data and include the planned R code analysis on that. We do that in all our Registered Reports, but we have it easy, because we use Qualtrics and that creates simulated data in seconds. This has a lot of advantages, in that it helps us all align, and sometimes reviewers can improve our code or catch mistakes. It also makes the plans more strict, so that we know exactly what we pre-registered to do. Vague descriptions of an analyses are very different from the specific planned code that runs it on simulated data.

b. I think in general Welch t-test is a good default, might make life simpler. (https://doi.org/10.5334/irsp.82)

c. Given the mixed literature, are you sure you want to run a one-sided test? In any case, please plan to report the effects and confidence intervals, and plot everything for us. A simulated data/code could help us get a better idea of what the final results section would look like.

d. Given the mixed literature, you might want to prepare for null effects. What would you do to quantify support for the null? Might be good to prepare a plan for equivalence/Bayesian in advance.

Response:

a. Thank you very much for this excellent suggestion! We have now simulated data using R and performed the analyses with this data set. We uploaded the corresponding R codes to OSF and added a results section to the manuscript based on this simulated data set (see p. 15 and following).

b. We have updated the manuscript to utilize Welch's t-test as suggested (footnote 11, p. 22).

c. We have thoroughly discussed this choice but decided to proceed with one-sided t-tests because our hypotheses clearly specify the direction of the expected effects. This choice is consistent with our theoretical framework and the specific predictions we have outlined. Nevertheless, we will report all effect sizes, confidence intervals, and provide detailed plots for clarity. (footnote 11, p. 22).
d. Thank you for your valuable feedback. We acknowledge the importance of preparing for null effects, especially given the mixed literature. However, our study is designed to test specific (directional) hypotheses rather than null hypotheses. Therefore, we have opted not to include equivalence/Bayesian tests at this stage. We may consider such analyses in our exploratory data analysis section if warranted.

Reviewer # 2 (Rajarshi Majumder):

15. First, I recommend minimising the elaborate presentation of the findings from past research. I understand the importance of emphasising the results of earlier studies, given the effects of identifiability are highly contested. In a potential resubmission, the authors should carefully re-evaluate whether the detailed explanation of the findings is a necessary part of the manuscript. This sometimes diverts the attention from the main arguments for the interaction hypothesis, and the readability suffers. It would be preferable to give this information in a concise tabular format.

    Response:
    Thank you for this valuable suggestion. We understand that the way previous studies and evidence are presented (for example, as text vs. a table) is, to some extent, also a matter of personal style and preference. Given the limited number of studies on the offender identifiability effect and the strong variation in study designs (and results) between the studies, we consider it relevant to highlight the (methodological) differences between the studies. We would, therefore, refrain from converting this information into a table. However, given your comment, we have streamlined the study descriptions to improve readability wherever possible, without omitting relevant aspects of the respective studies that are important from our perspective.

16. I had difficulty understanding the distinction between the current inquiry and the study conducted by Kogut (2011). In my opinion, the suggested inquiry appears to be a conceptual replication of Kogut's (2011). If this is indeed the case, it is essential for the authors to explicitly acknowledge this. Given the crucial aspect of the paper is examining the interaction effect between identifiability and the role of the punisher, it is imperative to discuss the need and implications of understanding (or potential revisiting) this effect.

    Response:

    We thank you for your feedback and acknowledge that our study may potentially be perceived as a conceptual replication of Kogut (2011). However, this also strongly depends on how you define the idea of “conceptual replication”. Below, we aim to

clarify the central aspects which, in our opinion, distinguish our research from a conceptual replication of Kogut's research:

Conceptual replications aim to test the same theoretical hypotheses using different methods or under slightly different conditions (Hudson, 2021). Kogut's research focused on comparing the perspectives of victims and offenders. In her second study, she investigated the role of emotions in the relationship between offender identifiability and punishment, where perspective did not play a role and people were not asked to put themselves in a specific perspective (hence, one may interpret these results as stemming from a third-party perspective). Importantly, Kogut did not formulate any theoretical assumptions or tested hypotheses on the effect of offender identifiability on the punishment intentions of uninvolved third parties, nor did she directly compare this perspective with the victim's perspective.

From a practical standpoint, this distinction is crucial because offenders (which were the focus in Kogut's research) do not usually punish themselves, and examining and comparing the offender's perspective cannot explain the mixed findings reported in previous research (as outlined in our theoretical introduction).

As described in our revised manuscript, our study fills this gap by directly comparing the victim and third-party perspectives in a well-powered pre-registered study. This approach seeks to provide more clarity on why previous research has yielded such divergent effects.

In sum, while we generally agree that Kogut's research is particularly relevant for the present study, we do not consider this a conceptual replication.

Hudson, R. (2023). Explicating exact versus conceptual replication. Erkenntnis, 88(6), 2493-2514.

17. The existing literature on the examination of the impacts of identifiability has several mixed findings. Although the authors cited the meta-analysis of Lee and Feeley (2016) as an indication of the identifiability effect, I will be cautious in interpreting the presence of a weak identified victim effect (r = .05) advantageous (with the three highest-powered studies in the dataset showing effects that are almost zero. For example, one study: 12,802 participants, r = 0.004). Further, a reanalysis of this meta-analysis by Maier et al. (2023) also uncovered moderate evidence of publication bias (BF01 = 0.11) and strong evidence of the absence of an identifiability effect. In addition, Vu et al. (2024) also found no effect of identifiability on altruism. Given the effect of identifiability is contested so much, I wonder if we need another research paper in this portfolio. – I ask the authors to stress what the investigation might add to the current literature, even if there are boundary effects to it or proves to be no significant difference between any of the proposed conditions in future. This is necessary to interpret in case of different outcomes.

Response:

First, thank you very much for your feedback, which echoes the comment made by Gilad Feldman (Reviewer 1) on our paragraph on the victim identifiability effect. As per our response to comment #7, we have strongly revised the paragraph on the victim identifiability effect to reflect the mixed results reported in the literature (p. 5, see the revised paragraph posted in our response to comment #1).

Additionally, we would like to emphasize the importance of the planned research in the context of existing literature:
Our study contributes by focusing on the identifiability of offenders rather than victims. This area has received less attention and, given the contrasting effects reported in the literature so far, a preregistered, well-powered study like ours is crucial to advance understanding. We aim to clarify how identifiability influences punishment judgments from both victim and third-party perspectives, offering insights into why previous research has shown divergent effects of offender identifiability.

Lastly, as per our response to comment #1, we would argue that especially when literature is so contested, it is even more important to conduct pre-registered research.

18. Is mediation analysis necessary? The feelings of empathy/moral outrage/blaming toward the offender should be correlated to the extent of punishment based on the identification levels [based on Study 1 of Kogut (2011) or similar analysis from Kogut & Ritov (2005)]. In my opinion, displaying the correlations should reliably predict the directions. In my view, the mediation hypothesis introduces unnecessary complexity in the paper.

    Response:

    Thank you! Your comment on the complexity of the mediation hypothesis is in line with comment #9 (by Gilad Feldman) and comment #4 (by Clare Conry-Murray). As detailed in our response to comment #4, we have decided to remove the mediation hypothesis and investigate possible mediations exploratively.

19. Furthermore, I failed to comprehend the justification for considering the variables as parallel mediators. It would be great if the authors could illustrate the reasons why the mediator residuals should not be correlated, as this can attenuate the extent of the effects of mediators on the dependent variable (paths b). While the authors referenced the mediation analysis using Preacher & Hayes (2004), it is important to note that collinearity could potentially impact the conclusion, as suggested by Preacher & Hayers (2004) in the same paper.

    Response:

    Thank you for your detailed feedback. Given that we have decided to remove the mediation hypothesis and analysis based on your recommendations, concerns

regarding parallel mediators and collinearity no longer apply to our current approach. However, we will certainly consider your helpful comments when conducting exploratory mediation analyses.

20. The study design template (page 22) is very difficult to read. The authors should consider making it precise and short. Additionally, I would suggest highlighting any potential post hoc analysis that the authors wish to investigate (for example, ANOVA controlling for empathy).

    Response:
    Thank you for this suggestion. We agree that our study design template can be more concise. We have revised the template to enhance clarity and precision, while retaining all critical information (p 25-26). As for potential post hoc analyses, such as controlling for empathy, we consider these as exploratory. Following PCI's guidelines to avoid including exploratory analyses in the Stage 1 report, we have not detailed these analyses in the current submission.