

## **Response letter (Stage 2, Round 1)**

Dear Dr. Fillon (Recommender), Dr. Celniker (Reviewer 1), and Dr. Kouassi (Reviewer 2),

Thank you for dedicating your time and expertise again to our project. We are very thankful for developing our manuscript with you and hope to have addressed all comments appropriately. As usual, we replied to them in the order of reception but grouped some of them together to avoid repetition.

We are looking forward to your impression of our updated manuscript and to the next steps for this project.

Best regards

Leopold Roth (corresponding author)

**Dr. Fillon (Recommender)**

Dear Dr. Fillon,

thank you very much for your continued support of our project.

We hope to have addressed all comments by the reviewers appropriately and in line with your expectations.

**Dr. Celniker (Reviewer 1)**

Dear Dr. Celniker,

Thank you again for providing us with your thematic expertise and nuanced readership of our manuscript.

We think the manuscript greatly benefited from your input and look forward to your opinion on the implementation. Right in the beginning, two comments were grouped together to avoid repetition. We responded to the rest in chronological order.

**Detailed Comments Dr. Celniker:**

***Comment:** There are some potential ceiling effects in the care context that are discussed in the manuscript, though as I note below, I think the authors can enhance their discussion of these effects. But I do not think this potential ceiling effect necessarily undermines the quality of these data, as the ceiling effect may be part of the phenomenon (i.e., contextual variation/limitations) that the authors are interested in. Beyond that issue, the data seems to be of sufficient quality to test the authors' hypotheses.*

*Corresponding comment to the topic:*

*pg 27 - Ceiling effects on moral perceptions did not preclude observations of effort moralization in studies using people running for charity as the stimuli, I believe. The presented results, with potential ceiling effects, seem different in this respect. I want the*

*authors to consider this a little more explicitly. In particular, it seems appropriate to discuss the Bigman & Tamir work: the high morality rating in the low-effort condition suggests that there was little room for effort to "amplify" the positive moral evaluations. But why did this happen in the care context but didn't in the charity context, etc? The authors get at this a bit, but I think a discussion of Bigman and Tamir in this section can help make this discussion more cohesive and valuable to readers, to help point them in the direction of the kind of work that's needed to address the open question that is being identified.*

**Response:** Thank you very much for sharing your perspective on this matter with us. We agree that the connection to prior work by you and your team and by Bigman and Tamir was critically missing from our discussion and that it is a valuable reference to better position our present findings within the literature. We included a dedicated paragraph to this in our updated discussion, in which we compare the empirical observations from your study on the effect of donations to fundraising marathon runners as well as the study of Bigman and Tamir (solving math problems which are rewarded by charity donations) with our present findings. Further, we highlight—to our impression—relevant differences between the scenarios on the behavioral level. From our perspective, the behavior in the care context (present project) is more directly attributable as ‘moral’, while the behavior in the marathon or the math scenarios is only moral due to the following charity donations. We argue that this behavior is less inherently interpretable as moral behavior and is hence subject to a stronger effort moralization effect. Please refer to pages 25.

**Comment:** *Yes, the exploratory analyses were justified and appropriate, though I think some of those findings can be better integrated in the General Discussion (e.g., lack of gender norm belief effects aligns with pre-registered analyses finding limited impact of gender on effort moralization).*

**Response:** Thank you for making this very productive suggestion. We added a respective paragraph, of which we hope that it connects the findings of the planned and the exploratory analysis in a logic and meaningful way. Please refer to pages 26.

**Comment:** *pg 13 - I think they mean "wage" instead of "salary" for the care worker*

**Response:** Thank you for your advice on this. We were not aware of this difference and adjusted it respectively.

**Comment:** *pg 15 - Are these really different studies? They are different samples, but that would be the same for any between-subjects study, which is why independent-samples tests would be deployed to compare between these groups. I think this is a single study, and the authors seem to acknowledge this by analyzing the data with ANOVAs that include context (work vs care) as a between-subjects factor. So even if the authors used separate survey links and Prolific ads to recruit participants, the intention was always to compare results across contexts, so I think it's more appropriate to say that this is a single study.*

**Response:** Thank you for sharing this observation with us. We discussed this point intensely between the team and decided to tease the results of the studies stronger apart, to underline the independent nature of the collected data. While it would likely be correct as well to speak of two samples within one project, we decided to stick with the terminology of two Studies.

Mostly, to highlight different motivations between the studies, as we conceptualized Study 1 to replicate the findings within an earlier established context and Study 2 to extend our inference into a novel context. We hope you follow our argumentation in light of the new structure of the results. Yet, both Studies remain closely related, and the exploratory tests are still reported within the same section.

**Comment:** *pg 15 - need to fill this in: "two separate samples at month/year [Stage 2]," seems like there is another one on pg 16. Need to fill in more of these details.*

**Response:** Thank you very much for spotting this. We corrected the points by adding the necessary information to the manuscript.

**Comment:** *pg 17 - there are some lapses in APA style throughout, eg "by Celniker et al. (2023) and Tissot & Roth (Tissot & Roth, 2025),". I personally don't care about this, but it will be important to address this to get the paper published.*

**Response:** Thank you very much for pointing this out to us. We corrected this mistake and checked the entire manuscript to ensure its compliance with APA style.

**Comment:** *pg 17 - I want to know more about the breakdowns for why participants were excluded. Was it mostly attention failures? Also, do the results hold when including the full sample of participants (as a type of robustness check)*

**Response:** Thank you very much for bringing this interesting question to our attention. As suggested, we have now added the exclusions by reason for each study in the manuscript (in "Data cleaning", p.16). Further, we added a robustness analysis to the main effects (Aim 1)

by study. In this analysis, we did not exclude participants who saw the high-effort condition as more effortful. We did not compute robustness checks for individuals who were excluded for other reasons (e.g., careless responding, missing language skills). While some effects changed in magnitude, most effects that were of meaningful size after applying the pre-registered criteria remained above the threshold when not applying all pre-registered criteria. The same applied to effects that were below our SESOI threshold ( $d = 0.20$ ). Interestingly, as also reflected in the discussion now, the low-effort person was now seen as more moral (in the value commitment dimension) in the care context.

**Comment:** pg 18. - *I want the authors to describe the results of Aim 1 more, not just show the table. Just a sentence or two about which effects replicated and which didn't.*

**Response:** Thank you for flagging this for us. We added a brief summary of the most relevant observations in the table. We hope it is now easier to comprehend and smoother to read. Please refer to pages 17-18.

**Comment:** pg 21-22 - *I would put a caveat on the comparisons of pay deservingness across contexts since it was measured on different response scales. I also don't think Figure 3 is necessary as it doesn't add clarity about the data beyond what's described in the text*

**Response:** Thank you for bringing this up. This also corresponds to the main argument by Dr. Kouassi (reviewer 2). We fully agree with these arguments and have modified the section strongly. We removed Figure 3 and separated the results strictly by Study. To do so, we adjusted the analysis from 2X2X2 ANOVA to two 2X2 ANOVAs. We hope these adjustments will clarify the results section and be more informative to future readers. Please refer to page 21.

**Comment:** pg 22 - *I would caution against saying the assigned cooperation result contradicts prior work. The assigned partner dynamic is very different. It's fairly logical (and consistent with past work) to want to choose a more moral/trustworthy partner (compared to a more competent partner) when the choice is available but to be more satisfied with the more competent partner if forced to work with them. And I think a discussion of whether these results contradict or just get at a different angle of cooperative dynamics is better saved for the discussion section. The authors do this to some degree in the GD, but I think they waffle a bit in terms of whether these findings contradict prior work or not. I think it doesn't contradict it, but wherever the authors land on this issue I think they need to talk about it consistently.*

**Response:** We appreciate this helpful comment and fully agree that the findings on cooperation satisfaction do not necessarily contradict prior work. We instead highlight that—corresponding to recent work—cooperation behavior depends on the situative affordances of a given task (e.g., trust- or competence-focused). Further, we emphasize potential effects of similarities versus discontinuities between task features (see also comment by Dr. Kouassi), since the present studies featured tasks in the same context as the original vignette (vignette: work context, cooperation type: work assignment), while earlier studies had a greater distinction between the vignette and the follow-up task (vignette: work context; cooperation type: trust game). In line with your suggestion, we have removed all phrasing that implied contradiction with previous findings and instead clarified in the discussion that these differences likely stem from methodological differences in how cooperation dynamics were studied. In addition, a paragraph was included in the limitation section to address the potential for differing outcomes resulting from choice versus assignment dynamics. Please refer to page 27-28.

**Comment:** pg 24- 25 - *"Hence, it appears more likely that gender role beliefs, as expressions of conservative views, moderate the effort moralization effect, yet to a very small extent." - it seems more accurate to say that gender role beliefs have a direct influence effort moralization, rather than having a moderating influence. If I'm understanding the results correctly, it seems like those with greater gender norm belief endorsement moralize effort more (i.e., display larger differences in value commitment evaluations between the low & high effort targets). I'd ask the authors to spell out the results a little more slowly and carefully to make sure the reader comes away with the correct understanding of these results.*

**Response:** Thank you very much for highlighting this error in our reporting. We corrected the interpretation of the analysis, which is hopefully better to understand in its current form.

Please refer to page 26.

**Comment:** pg 25 - *again, I would like the authors to slow down and explain the participant gender by target gender interactions so that I can make sure I understand the effects. I don't have a clear understanding of what the authors found with the results as currently written. It seems like generally null effects, but I don't understand the the details of the significant effect that was detailed.*

**Response:** Thank you for sharing this with us. We added some further explanation to this analysis, hoping that it will clarify the results. Further, we reiterated the fact that this finding holds no practical relevance, given its effect size. Please refer to page 26.



**Comment:** pg 26 - the authors talk about the differences in cooperation findings better here. still use language of contradiction though, which goes against what they are arguing in this paragraph. And maybe a more pointed statement is warranted asking for future research to investigate when the influence of moral perceptions and competence perceptions in cooperative dynamics?

**Response:** Thank you for the comment. Please refer to our response on the collaboration satisfaction findings above.

**Comment:** pg 27-28 "The only gender-related finding was that, in the work context, female actors were preferred as cooperation partners over male actors, irrespective of effort level." Were there vignettes where a woman and man were each presented? That didn't seem like it was the case from the methods. If not, I don't think it's right to say that they were "preferred" over male actors, since participants only ever chose a woman over a woman or a man over a man. They make this clear on pg 30, that woman and men were never compared by the same participant. So I would rephrase this to make the point clearer to readers.

**Response:** Thank you very much for this important comment. We fully agree that this section was not optimally worded and 1. corrected it, by highlighting the limitations of this findings and 2. shortening it, to increase readability. We hope the description is now appropriately nuanced. Please refer to page 26.

**Comment:** pg. 28 - It might be clearer to say "Overall, these findings suggest that the moralization of effort applies similarly **to judgments** of men and women.."

**Response:** Thank you for this suggestion. We chose something very close to this formulation. We hope the summary of these (absent) observations is appropriate and usable for dissemination. Please refer to page 26.

***Comment:** In general - There are ways to trim the materials, procedure, and results sections moving forward. I appreciate that the authors kept the manuscript structure the same for this Stage 2 report, which made it easier for me to review. But I think they will have an easier time getting this paper published after it goes through PCI if they refine these sections a bit more. This is not important to me to get this Stage 2 protocol approved, but I just wanted to share this feedback to help the authors as they move toward getting this piece out into the world*

**Response:** Thank you very much for this suggestion. We took this very seriously and tried our best to trim the manuscript wherever we detected lengthiness. Given the extended content of the discussion, the article is not meaningfully shorter by now, but the information density should be higher and the communication of our findings clearer, compared to the prior version.

**Dr. Kouassi (Reviewer 2)**

Dear Dr. Kouassi,

Thank you very much for your productive and positive feedback on our manuscript. We greatly benefited from your input and hope to have included everything in line with your expectations.

**Detailed Comments Dr. Kouassi:**

***Comment:** The arguments presented concerning the moral dimension and the differences in terms of intrinsic motivations of the “work” and “caregiving” contexts are, in my opinion, very relevant in giving readers avenues of exploration for future research on the differences in moral judgment between these two contexts. For example, as mentioned in the article, if participants feel that caregiving is done while being paid less, it's because they probably feel that other (perhaps more intrinsic) motivations drive them to perform these tasks. This could in fact have an influence on participants' perception of these different protagonists and thus explain the differences between the “work” and “caregiving” contexts.*

*The results showing a suggested difference in wages between the “work” and “caregiving” contexts do indeed support these arguments. However, in my opinion, it is very risky to present these results as they stand.*

*The limitation you mentioned regarding the results showing a difference in the wages awarded between the “caregiving” and “work” contexts is indeed very true. The reference salary may well have had a strong influence on the salary awarded in these two contexts, and this constitutes a confounded effect which prevents us from correctly concluding that there is a difference in responses between the “work” and “caregiving” contexts. It should also be pointed out that including this “context” variable (work vs. caregiving) in the suggested*

*wage predictor model is not consistent with the fact that the 2 studies are presented as independent and distinct, nor with the fact that the other models tested to predict the other VDs do not include this variable. Moreover, the use of this variable was (unless I'm mistaken) not presented in Stage 1.*

*Insofar as the conclusions that can be drawn about the context variable (work vs. caregiving) are limited by the existence of a confounded variable, and that it breaks the coherence of other components of the item, is it still relevant to present these results and conclude on them?*

*Even though these results would be interesting to illustrate, I recommend that the authors remove them so as not to overcomplicate the article and mislead the reader.*

*In the “limitations” section, it may be pointed out that, due to the design of this work as two separate studies, the comparison between the two contexts has been rendered irrelevant, and that future work is recommended if these two contexts are to be compared.*

**Response:** Thank you for raising these important methodological concerns. We decided to remove all direct comparisons between the work and care contexts from our analyses and findings regarding pay deservingness, thereby removing the confounding variable of 'context' from the analyses altogether. We no longer interpret differences between contexts as meaningful findings. We have instead added a short explanatory note in the limitations section that clearly acknowledges the methodological differences between the studies—i.e. the presence or absence of a reference salary/wage—and highlights the likelihood that anchoring effects have influenced these results. Please refer to page 28.

**Comment:** *With regard to the unexpected findings on partner choice, it is highly relevant to mention the difference in context with previous studies: Celniker et al. (2023) did indeed focus on partner choice in the context of a game, whereas the present studies focus on partner choice in a context of cooperation in the world of work or caregiving. In order to guide future studies on the subject to possible explanations for these differences, it may also be relevant to mention another difference between the 2 studies. This difference lies in whether or not the measure of partner choice refers to the same task on which these potential partners are presented. In the present studies, the protagonists are presented as making efforts or not in the same domain as that on which the participants will be led to imagine themselves cooperating (e.g., a cooperation task that remains affiliated with the world of work). This is not the case in Celniker et al. (2023), where the partner is presented as doing or not doing a lot of effort in sports tasks (marathons), but where the choice of partner is measured in a “trust game” task that does not belong to the same domain as marathons. This difference in protocol may be relevant to the discussion.*

**Response:** Thank you for highlighting this important methodological distinction. Following your suggestion, we have now explicitly addressed the difference between our approach and that of Celniker et al. (2023) in the discussion section. Please refer to pages 27-28.

**Comment:** *In the participants section, it might be useful to specify information on the socio-economic status of participants, as well as their level of education (these data are collected in the questionnaire).*

**Response:** Thank you very much for raising this critical point. We added the information to the sample description. Please refer to page 12.

**Comment:** *At page 8, in “This observation was replicated well, yet it appears to vary between cultures in magnitude (Mexico:  $d = .14-.28$ , Germany:  $d = .34-.37$ , France:  $d = .38$ , US:  $d = .60$ , South Korea:  $d = .71$ , (Celniker et al., 2023; Tissot & Roth, 2025).”*

*There's an extra parenthesis just before “Celniker”.*

**Response:** Thank you very much for your attentive readership. We corrected the typo in the updated manuscript.