I appreciated the opportunity to review this Stage 2 manuscript. Overall, the authors conducted the work they said they were going to, so I think they upheld their end of the bargain. That said, I think there are several edits that can be made to make the manuscript more informative and more easily readable that the authors should consider. I will first cover the main PCI criteria for Stage 2 manuscripts before discussing some of my more minor feedback in greater detail.

**2A. Whether the data are able to test the authors’ proposed hypotheses (or answer the proposed research question) by passing the approved outcome-neutral criteria, such as absence of floor and ceiling effects or success of positive controls or other quality checks.**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.

**2B. Whether the introduction, rationale and stated hypotheses (where applicable) are the same as the approved Stage 1 submission.**Yes, this report was consistent with the State 1 submission.

**2C. Whether the authors adhered precisely to the registered study procedures.**Yes, the authors adhered to the Stage 1 protocol and procedures.

**2D. Where applicable, whether any unregistered exploratory analyses are justified, methodologically sound, and informative. Yes, they 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).**

**2E. Whether the authors’ conclusions are justified given the evidence. In general, I think the authors’ conclusions are justified, though as I note below, I think there are some places where they should be careful or take more time to explain their perspective (e.g., discussion of whether or not the cooperation results “contradict” past work).**

Here are some more minor points and elaborations on some of the issues discussed above:

* pg 13 - I think they mean "wage" instead of "salary" for the care worker
* 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.
* 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.
* 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.
* 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
* 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.
* 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
* 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.
* 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.
* 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.
* 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?
* 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.
* 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 enver compared y the same participant. So I would rephrase this to make the point clearer to readers.
* 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.."
* 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