**0. General advice**

The paper addresses an interesting question, focusing on the “effort moralization effect”. The authors announce their intention to replicate the effect in two areas: work and caregiving context. At the same time, they raise the question of whether the effort moralization effect is moderated by gender, by presenting the various predictive models available in the literature on the subject. Overall, the introduction sets out the subject well and supports its arguments correctly. The two studies presented are coherent with what the authors claim to demonstrate, and the methodology makes it possible to establish a relevant cause-and-effect relationship. The fact that the different possible predictions based on different models of moral judgment are considered is a pertinent point and enables readers to clearly visualize what can be expected from these studies. However, we have identified a number of points to bear in mind in order to optimize the registered report.

**1. Major issues**

***1.1. Model A: The graph could be adjusted differently in model A.***

If we expect women to be punished more than men in the moral judgment made of them for violating expectations, then shouldn't women in the low-effort condition be judged more negatively on their morality than men in the low-effort condition? According to the results of Kennedy et al. (2016), women were indeed judged more harshly than men in the event of unethical action. Indeed, in the present studies, why would women who make no effort (and therefore violate observers' expectations) remain better judged than men (including those who make little effort) in these studies?

***1.2. Models and hypotheses***

The different models presented (A, B, C and D) give readers a clear idea of the results that can be expected from the two studies. To optimize and clarify this understanding, it might be appropriate for the authors to include a paragraph that clearly indicates the situation in which we might expect to see the different models expressed. For example (this is just an example, and not necessarily what needs to be stated), mention the fact that, according to the literature, one might expect one model (e.g., model A, B or C) to express itself in the world-of-work study, while another model (e.g., model C or D) will express itself more in the case of caregiving, and/or explain which models are the most credible according to the literature in each situation. This could give greater visibility and clarity to the hypotheses.

***1.3. Relational aspect in caregiving***

The tasks described in the care condition are all non-relational tasks (cleaning, preparing meals, etc., rather than spending time with the person, chatting, taking them for a walk, etc.). However, more relational tasks would correspond to an important aspect of caregiving, would theoretically be closely linked to the warmth dimension and could potentially be judged as a greater emotional investment (similar to studies by Johnson & Park, 2021). Indeed, including these relational tasks would better illustrate the effects insofar as caregiving tasks could be more accurately categorized as warm.

Depending on the research objectives, it might be relevant to explain why the relational aspect was not considered in the conditions of the caregiving study. And if research objectives allow, it might even be possible to include more relational tasks in the text of the caregiving study conditions.

***1.4. Sample size***

In line with the recommendations of various works on power analyses of interaction effects (cf. this link: <https://approachingblog.wordpress.com/2018/01/24/powering-your-interaction-2/> and see also Sommet et al., 2023: <https://doi.org/10.1177/25152459231178728> and this app [www.intxpower.com](http://www.intxpower.com)), the G\*Power analysis that has been carried out by the authors, although it seems appropriate for the main effects targeted, appears to be insufficient for analyzing the power of interaction effects (e.g., the interaction between gender and amount of effort produced). I therefore recommend that the authors include in the document another power analysis that takes into account the specificities of an interaction effect (using, for example, the resources proposed above or any other resource that would be relevant for a power analysis of an interaction effect). In view of the effect sizes targeted by the registered report (small), the sample size required is likely to be much larger than initially planned (initially anticipated at N= 350 per study according to the G\*Power analysis currently posted). In fact, the authors will probably have to choose between these two options:

1) Increase the sample size of the two studies to bring them into line with the recommendations on the detection of interaction effects. This would be the most desirable option, as it would maintain the ambition of the paper and thus provide solid support for the literature. However, it is also very costly, and there is a strong possibility that, due to various constraints, it will be impossible for the authors to obtain a sufficient sample.

2) Indicate in the document that it was not possible to recruit more participants, and that the sample as presented is indeed too small in relation to the recommendations made for the detection of an interaction effect. The ambition of the registered report will therefore be scaled back, but it will still provide a contribution to the current literature.

**2. Minor issues**

***2.1. Introduction: Female high morality model (A): typing error***

There are two spaces after the sentence “The effort shown at work might be perceived as a manifestation of these stereotype dimensions.”

***2.2. Procedure: typing error***

In the 3rd line, a period is missing at the end of the sentence: “The data was collected in two separate samples at month/year [Stage 2], with participants from one study being excluded from participating in the other”

**3. Conclusion**

The article makes an interesting contribution to the literature. In particular, it extends the “effort moralization effect” to other contexts (caregiving) and proposes a legitimate and natural link with gender differences and the associated differences in judgments. The means employed (in terms of sample size) may perhaps curb the ambitions of these studies, but the fact remains that the document proposes interesting answers to the questions it poses, and its form allows a good understanding of the subject. On a personal note, I would recommend a resubmission and adjustments by the authors with a view to future publication.