**Geneva, March 27th 2024**

**Review of PCI RR**

**Is it Worth the Hustle? A Multi-Country Replication of the Moralization of Effort Effect and an Extension to the Increasing Aversion to Bullshit Jobs**

This paper aims to partially replicate and extend Celniker et al. (2023) Study 6, which tests the effort moralization effect (i.e., the notion that ceteris paribus, people who are exerting more effort in a job are seen as more moral) in countries that have not been tested in the original, i.e., Germany, Mexico, Netherlands, and South Africa. The extension aims to test the effect of age on the effort moralization effect. Overall, I think there is merit in this idea, but there is a lot of effort to be put in before the study can be conducted.

1. Writing
   1. In general, I think the writing in this paper might be more to the point. For instance, rather than starting with Aristotle, I would prefer you start with a description of what you want to do in the paper. The reader of scientific papers wants to know what you do and what you find, not read introductions.
   2. There are several sections that seem redundant and not in line with the topic, for instance the section about faces. Your paper is not about faces, so I do not think this section should be included. Similarly, the section about phrenology, racism and morality inferences is out of topic. Your paper is not about phrenology or racism. I would replace it with a section about the many cues that lead people to infer morality, for instance.
   3. The main effect in the abstract should be explained in a clearer way. Right now, it reads ‘The effort moralization effect describes the process of deriving information through the effort invested into a given task’. I would change it to ‘The effort moralization effect is the finding that people who are exerting more effort in a job are seen as more moral’.
   4. Personally, I do not think that all the discussion of quiet quitting – a concept I find ill-defined – is warranted. If you want to test the effect of age on the EME in your paper, that is fine, justify it with scientific literature.
   5. When writing, you are using too many commas. For instance, in the sentence at p.9 ‘One phenomenon, that has raised scientific psychology’s interest in recent years is the observation, that people appear to use effort, invested in given tasks, as information on the morality of agents, further summarized as effort moralization effect.’ There should be no commas.
   6. At p. 10, you write ‘building on this finding, Celniker et al. do XYZ’, but you have already described Celniker et al. in the previous paragraph.
   7. You often combine together Bigman and Tamir 2016 and Celniker et al. 2023 as both showing an effort moralization effect, but that is not really correct. B&T look at the amplifying effect of effort. If you regard Celniker et al. 2023 as part of the findings of B&T (because it only looks at positive outcomes perhaps) then you should more clearly explain so.
2. OSF
   1. Looking at the word documents of the survey, they still have a default placeholder in the informed consent page. See below for an example in the case of the German survey. Please double-check everything. Also note that the OSF file viewer has issue with Word (it shows the files as having many more pages than they actually have). I recommend you upload pdf files instead.

A screenshot of a computer screen

Description automatically generated

1. Methods
   1. You should not apply a within-test alpha correction for the number of tests. The tests you want to do are paired-samples t-tests (you call them dependent t-tests, which is fine) but they are all independent from each other. This analysis is not an ANOVA with multiple pairwise comparisons where the pairwise comparisons tests are dependent on the main analysis.
   2. It is not clear to me how you translated the original text of the survey in Celniker et al 2023 into several different languages. Please document the translation in much more detail. This is a crucial passage and cannot be liquidated in a couple of sentences and a citation as you do right now.
   3. You have a series of completed data in several tables (e.g., Table 6 and 7). Are they part of a pretest? Are they randomly generated data (as I suspect)? You need to specify it.
   4. Why did you choose Germany, Mexico, NL, and SA? Convenience sampling or other reasoning? This is also an aspect you should describe in further detail.
   5. In your survey, there are questions related to participants’ future of work. These questions are barely introduced in the paper. What is the point of these questions? Where do they come from? Are they validated? How will they be analyzed? Please specify.
   6. See above for the status ladder. Please specify why you include each measure and how do you intend to analyze it.

I have other comments that follow the PCI RR guidelines, importantly about how you plan to interpret the results – please see below in bold.

* Does the research question make sense in light of the theory or applications? **Yes**
* Is it clearly defined? **Yes but more detail is needed in several places.**
* Where the proposal includes hypotheses, are the hypotheses capable of answering the research question? **Yes**
* Is the protocol sufficiently detailed to enable replication by an expert in the field, and to close off sources of undisclosed procedural or analytic flexibility? **Not yet. More detail is needed, see above.**
* Is there an exact mapping between the theory, hypotheses, sampling plan (e.g. power analysis, where applicable), preregistered statistical tests, and possible interpretations given different outcomes? **No details are given regarding the interpretation if the findings are not conclusive or if the tests are not statistically significant, and more is needed.**
* For proposals that test hypotheses, have the authors explained precisely which outcomes will confirm or disconfirm their predictions? **No, please see above.**
* Is the sample size sufficient to provide informative results? Not so much.
* Where the proposal involves statistical hypothesis testing, does the sampling plan for each hypothesis propose a realistic and well justified estimate of the effect size? **It proposes the one of the original paper, but with 99% power. I do not think that the use of Bonferroni correction is warranted, so overall I think that the sample size that authors propose is adequate.**
* Have the authors avoided the common pitfall of relying on conventional null hypothesis significance testing to conclude evidence of absence from null results? **Yes**
* Where the authors intend to interpret a negative result as evidence that an effect is absent, have authors proposed an inferential method that is capable of drawing such a conclusion, such as [Bayesian hypothesis testing](https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00781/full) or [frequentist equivalence testing](https://journals.sagepub.com/doi/full/10.1177/2515245918770963)? **Unfortunately not, and I encourage the authors to do so. In general, this being a replication, you should use the LeBel et al. (2019) guide to both a) the methodological facets of the replication and b) the quantitative comparison between the original and the replication results. At p. 20 you should also specify the coverage of both confidence and credible intervals (e.g., 95%).**
* Have the authors minimised all discussion of post hoc exploratory analyses, apart from those that must be explained to justify specific design features? Maintaining this clear distinction at Stage 1 can prevent exploratory analyses at Stage 2 being inadvertently presented as pre-planned. **Yes, it seems so.**
* Have the authors clearly distinguished work that has already been done (e.g. preliminary studies and data analyses) from work yet to be done? **Unsure, see my comment above on Tabe 6 and 7.**
* Have the authors prespecified positive controls, manipulation checks or other data quality checks? **Yes, because they follow the original paper.**
  + If not, have they justified why such tests are either infeasible or unnecessary? Is the design sufficiently well controlled in all other respects? **NA**
  + When proposing positive controls or other data quality checks that rely on inferential testing, have the authors included a statistical sampling plan that is sufficient in terms of statistical power or evidential strength? **Yes.**
* Does the proposed research fall within established ethical norms for its field? Regardless of whether the study has received ethical approval, have the authors adequately considered any ethical risks of the research? **Yes**

Overall, I think this idea is good, but there is some work to do.

I know that I made a lot of observations, but it seems to me that this paper needs to be refined quite a bit.

Best

Ignazio Ziano (University of Geneva) – [Ignazio.ziano@unige.ch](mailto:Ignazio.ziano@unige.ch)

**References**

LeBel, E. P., Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A brief guide to evaluate replications. *Meta-Psychology*, *3*.