

Response to Reviewers

Reviews (Round 2)

Dear Dr. Rahal,

we thank you for the detailed feedback and comments on our manuscript submission. Here, we provide a point-by-point response to your main points and each of the reviewer's comments. In particular, we have revised the instructions based on R2's feedback and provided more information with regard to hypotheses and sample size rationale based on R3's comments. We hope that you will consider this revised manuscript for Stage 1 acceptance.

Editor

I have now received three re-reviews. There are few remaining issues to integrate:

- clarify experimental procedures and instructions

We have adapted the final instructions based on R2's feedback.

- clarify the sample size rationale

We have further clarified the power analysis and sample size rationale based on R3's feedback and now highlight that the power analysis for H1 refers to total $N = 600$. Detailed response to each reviewer are provided below.

R1: Review by Felipe Vilanova, 14 Apr 2025 18:44

I think that the authors sufficiently addressed my concerns and the registered report is ready for data collection.

R2: Review by anonymous reviewer 1, 23 Apr 2025 08:56

I thank the authors for the detailed response to my previous comments. According to my assessment the revised manuscript indeed represents a much better aligned research design. I

believe all of my previous concerns have been sufficiently addressed in the revision. In the following, I would only like to suggest some minor changes to the experimental instructions.

To clarify what happens each round and what is based on all decisions, I would suggest the following general changes:

“2. Company’s Income and Income Taxation” Change this to “2. Company’s Income and Income Taxation **each round**”

“3. Company Success” Change this to “3. Company Success **each round**”

“3. Audit of the Reported Income” Change this to “3. Audit of the Reported Income **each round**”

Alternatively, you could also enter an introductory sentence to clarify that income and success calculations happen on a round-by-round basis just to be sure that participants do not mix this up as they only receive certain information in the end.

We thank you for the suggestion and have added “each round” for each of the three headings across all treatments.

I further have some minor wording suggestions in the specific treatment instructions to enhance clarity:

2.2 Instructions Gr-No & Gr-Ind & Ind-Gr & Gr-Gr

“You will only receive information **on the other employee's report** at the end of the experiment.” Change this to: “You will only receive information **on the other employees' reports** at the end of the experiment”

“The tax will be calculated based **on the average report of the income** that you and your two colleagues report” Change this to: “The tax will be calculated **based on the average of the incomes** that you and your two colleagues report”

We have changed both descriptions accordingly.

2.3 Instructions Ind-Ind

“The company success is 1000 for a tax of 0 (reported income 0). If audited the unpaid tax needs to be paid (250) and a fine of the same amount (250) resulting in a company success of 500 (1000 -250 -250).” Change this to “The company success is 1000 for a tax of 0 (reported income 0) **if your report is not audited**. If audited the unpaid tax needs to be paid (250) and a fine of the same amount (250) resulting in a company success of 500 (1000 -250 -250).”

2.4 Instructions Gr – Ind & Gr-Gr

Could it be that in each of these instructions the example provided under “4. Company success” is missing information on the audit?

We have added the information that the first part applies if the report is not audited. We also added information on the audit for the group payment structure treatments. For example, for the Gr-Gr treatment it now adds:

“The company success is 1000 for a tax of 0 (reported income 0) if the group report is not audited. If audited the unpaid tax needs to be paid (250) and a fine of the same amount (250) resulting in a company success of 500 (1000 - 250 - 250) for your group.” (p. 21 Supplementary Material)

R3

I have read the revised stage 1 report as well as the response letter. Overall, I think that the authors addressed my comments sufficiently. I only have a few remaining minor issues. Please see below my detailed comments.

I thank the authors for their engagement with my comments and their work revising the manuscript. However, I do have a few minor comments on the revised manuscript. Below please find my specific comments.

Comments

1. Page 3, end of the first paragraph, should be “highly context dependent”

We have changed this typo.

2. Page 21 and 22, the description of H4 and H5 seem to be the same. If my understanding of the design Table is correct, H4 is focusing on the main effect while H5 is focusing on the no-interaction effect. Perhaps these points can be made even clearer for the readers.

We agree that this could be confusing and have added more information when introducing each of the hypotheses. As correctly interpreted, H4 focuses on the main effect of the type of punishment (i.e., individual punishment shows stronger reductions in dishonesty compared to group punishment) and H5 focuses on the interaction effect (i.e., H4 will hold regardless of payoff structure). H5 basically assumes a similar effect size for H4 for both individual payoff and group payoff.

“When looking at the predictions for each type of punishment across payoff structure, the expected gain from (full) dishonesty is lower for the individual punishment treatments compared to the group punishment treatments (Table 1). Given that there is limited empirical evidence on the effectiveness of group punishment (Bonfim & Silva, 2019) we predict a main effect of punishment with

individual punishment being on average (across individual and group payoff structure) more effective in reducing dishonesty compared to group punishment:

H4. Individual punishment (vs. group punishment) will show stronger effects in reducing dishonesty across payoff structure.

When considering the interaction of type of punishment and payoff structure we would expect a similar effect of individual punishment for both types of payoff structure. Based on the predictions (Table 1), individual punishment should reduce dishonesty more strongly compared to group punishment and to a similar degree for both individual payoff structure and group payoff structure. Therefore, we expect no interaction effect between type of punishment and payoff structure:

H5. Individual punishment is expected to reduce dishonesty more strongly compared to group punishment across both types of payoff structure.” (pp. 21-22)

3. The authors explained that for H1, the power analysis is done for a subsample. However, the main text still reads “In addition, we simulated a model accounting for our smallest effect size of interest when testing main effects in H1 and observed between 86% and 90% power at around 600 participants”. This can also be made clearer that 600 is referring to the N total and not the subsample sample for the analysis of H1.

We have added this information to clarify that this is referring to the total sample size:

“In addition, we simulated a model accounting for our smallest effect size of interest when testing main effects in H1 and observed between 86% and 90% power at around 600 total participants and around $n = 100$ participants per treatment (Supplementary Figure 2).“ (p. 24)

4. Page 27, “For instance, if participant A reports a company income of 1000, participant B a company income of 500, and participant C a company income of 750, then 750 will be selected

as the group-based reported company income. Each group member will therefore receive an individual bonus payoff of £1 ($812.5 = 1000 - 25\% \times 750$). Bonus payments per round”

It will be clearer to say “a mean of 750 will be” or something similar. The second highlighted part seems to be incomplete.

We have added “a mean of” to the specific sentence. We have deleted the second part as it seemed to be part of a previous draft that was not moved during editing.

5. Regarding the analyses of moral emotions and stress, I disagree with the authors regarding the second analysis: emotions -> compliance/ dishonesty. Since the items are asking about emotions given one’s own or peers’ behavior, I believe that the direction should be compliance/ dishonesty -> emotions. Even though, as the authors stated, the current dataset cannot assess causation, I believe that the direction of the regression model, which implies a causal direction, still matters.

The main focus of the analyses for moral emotions and stress is on differences in ratings across the experimental treatments. The last analysis is mainly added to test the relationship between moral emotions and compliance and as the reviewer highlights, focuses on a correlational association. We have now changed the direction and suggest three multilevel models with moral anger, guilt, or stress as the outcome and compliance as the predictor:

“Further, we will conduct three multilevel regressions with moral anger, guilt, or stress as the dependent variable and compliance (mean centered) as the predictor.” (p. 32)

6. Design Table, manipulation check, it should be clearer that we are referring to the rejection of the null hypothesis.

We have made it clearer that this rejection refers to the null hypothesis by adding the term.

“The null hypothesis is rejected if the pairwise comparisons with the group treatments (Gr-No, Gr-Ind, Ind-Gr, Gr-Gr) and the baseline are statistically significant, the 95% CI is not entirely included in the equivalence bounds, and social commitment is higher for group treatments compared to the baseline.

The null hypothesis is not rejected if the 90% TOST CI of the pairwise comparison is entirely included in the equivalence bounds.

If the null hypothesis is rejected the manipulation is considered successful.

If the null hypothesis is not rejected the manipulation is considered unsuccessful and we will investigate which treatments failed to increase feelings of commitment.” (Table 2)