PCI Registered Reports #730: Managing Disclosure Outcomes in Intelligence Interviews

Dear Zoltan,

Many thanks for inviting a revision of the captioned registered report. We have completed the revision and responded to the comments raised by the editorial team. Our responses appear in red font below. The revisions feature in TRACK CHANGES in the manuscript.

Sincerely
Authors

Recommender

Two reviewers have evaluated your stage 2 and are largely happy. They have some minor concerns to address. I just have one further point to make. In the manuscript you point out you have low power to pick up an effect for the crucial interaction of half the size found in the Pilot. As half the size may still be interesting, you point out that a non-significant result means one should suspend judgment. In the Discussion you point this out again. That is good. You also claim in the Abstract and Discussion that the interaction failed to replicate. On the one hand, if replication means "getting a significant result", that is true. But many people implicitly take the claim that an effect failed to replicate to mean there was evidence against the effect. Be clear in the Abstract that by failing to replciate you mean there was a significant interaction in the pilot but but not a signfiicant interaction in the main study, though the power means interesting effect sizes may have been missed.

Thank you for drawing our attention to the possibility of misrepresenting the meaning of a failure to replicate. We have now rewritten that aspect of the results in the Abstract and General Discussion sections.

Yikang Zhang

Managing Disclosure Outcomes in Intelligence Interviews is a stage-2 registered report focusing on empirically testing the proposed model of disclosure decisions in intelligence interviews. In their preliminary study, the authors found that unguarded information with low cost and high benefit were most likely to be disclosed while guarded information (high cost and low benefit) were most unlikely to be disclosed. Rates of disclosure for low-stake (low cost and low benefit) and high-stake (high cost and high benefit) information fell in between. In the planned study, the authors reported failure to find support for the Cost*Benefit interaction (an average effect in GLMM) found in the preliminary study. The authors speculate that this inconsistency resulted from the heterogeneity of participants' risk and benefit perceptions.

I have only a few comments on the stage 2 report:
1. Both in the analyses of the preliminary study and the replication, the authors reported the ICCs of unconditional model (referring to the models without fixed effects I think) and consider the ICCs of items represents the individual differences in participants' risk/ benefit perception of the item contents. In my opinion, this is only valid, if for every item, it can be paired with low stake/ high stake/ guarded/ unguarded based on probabilities. Based on my reading, this seems to not be the case. Instead, the items were fixed to be in certain conditions but not others. Perhaps it would be more appropriate to report the ICCs of items in the conditional models (which controls the experimental manipulation by including them as fixed effects) as the individual variation in risk/benefit perception of item contents.

We think this is an interesting point. We have added ICCs based on an additional model which adds the cost and benefit fixed effects (with no random slopes, as this makes ICC calculation problematic). These ICCs also demonstrated substantial variance for participants, ICC = .15, and for items, ICC = .43. Though the estimates differed from those in the fully unconditional model, we don't think they substantively change our interpretation of the results. These new ICCs are reported on p.29.

We also calculated a similar set of ICCs for a conditional model for the preliminary study. This analysis indicated considerable variance associated with participants, $ICC = .150$, and with items of information, $ICC = .486$. These results are very similar to the conditional model results of the main study. These results are presented on p.18.

2.	The authors seem to attribute to the non-significant result of the replication study to the individual variations in the perception of the contents (resulting in weak manipulation/ larger noise). I find this a valid explanation as well. However, this does not seem to explain the fact that with a similar 'weak manipulation', there was a significant result in the preliminary study. Moreover, the ICC of items reported in the unconditional model was larger than in the replication study, which could suggest a great individual variation in the former case (or conversely, a stronger manipulation, given we couldn't separate these two in an unconditional model, see my point 1).

Thanks for this comment. We agree that there is substantial individual-level variation. However, we do not see the manipulation in the replication study as weaker or even stronger. In truth, we remain agnostic about the origin of the difference between the Preliminary and the Replication Studies. As noted on page 31 (paragraph 2), "Our best speculation, at this time, is that our efforts to control variance stemming from the background story's narrative content did not work. Participants in the Replication Study, like in the Preliminary Study, may have construed their own estimates of cost and benefit probabilities based on the narrative content (as well as other idiosyncratic considerations), not minding the numerical manipulations." So, we think that it might be challenging to control such variation in any experiment. For that reason, we recommend that researchers study such individual variation rather than try to control it, at least until we better understand what explains this variation.

3.	If we look at the descriptive stats in the replication study, it is still consistent with the model's prediction, which could be of value to mention in the discussion.

Thanks for the reminder. We have included a sentence in the Replication Study's discussion mentioning that the disclosure rates (of the respective information-types) were largely in line with our predictions.

4.	Also, it seems that compared to the preliminary study, the differences in disclosure rates between conditions was smaller in the replication study. Perhaps the authors could compare the two studies and discuss what could be the reasons behind this.

5.	One way to avoid the large individual variation in risk/benefit perception is to measure participants' risk and benefit perception of these items directly and use these as predictors in the model. Of course, then we are not directly manipulating risk/benefit (even if we manipulate the factors, the ratings are a product of manipulation by trait interaction), thus losing 'power' for causal inference. Nevertheless, it could be a complementary approach to the current design.
For example, in one of my previous paper examining memory verification strategy (Zhang et al., 2023, also see Nash et al., 2017), we asked participants to rate the perceived cost and benefit of each strategy and use these ratings to predict the likelihood of using such strategy. If the authors agree with my point, perhaps this could be added in the discussion section.
(the authors need not to cite either my paper or the paper of Nash et al., 2017)
Reference
Nash, R. A., Wade, K. A., Garry, M., & Adelman, J. S. (2017). A robust preference for cheap-and-easy strategies over reliable strategies when verifying personal memories. Memory, 25(7), 890-899.
Zhang, Y., Nash, R. A., & Otgaar, H. (2023). Preference for cheap-and-easy memory verification strategies is strongest among people with high memory distrust. Memory, 31(7), 978-988.

Because these two comments suggest pathways for future research, we will address them jointly.
Many thanks for these suggestions. However, we prefer to stick with the analysis we preregistered and avoid too much speculation (and additional analysis), which leaves room for possible mis-citation and misrepresentation. Our plan is to explore (a) possible reasons behind the differences in the studies and (b) alternative ways to examine cost and benefit perceptions in future work where we can subject our speculations to preregistered testing.

6.	For the concluding remarks, perhaps it could be useful for readers if (in)consistent results from the current study are briefly mentioned as well, instead of only 'what is needed to be done in the future'.

Given that we have mentioned the inconsistent results in the abstract, the replication study's discussion, and the general discussion, we prefer to leave the conclusion concise to make the manuscript concise. Suppose the recommender advises we add such a note to the conclusion; we will comply.

I hope that the authors find the my comments helpful and I apologize in advance if I misunderstood the MS in my reading.

Thank you for the great comments and suggestions for future research. We intend to incorporate your suggestions in our upcoming research, particularly the suggestion to examine subjective cost and benefit perceptions.

Tyle Jacobs

Overall, I am satisfied with the authors' reporting of their Stage 2 results. PCI-RR highlights the following questions as important for reviewers to consider at Stage 2: Have the authors provided a direct URL to the approved protocol in the Stage 2 manuscript? Did they stay true to their protocol? Are any deviations from protocol clearly justified and fully documented?Is the Introduction in the Stage 1 manuscript (including hypotheses) the same as in the Stage 2 manuscript? Are any changes transparently flagged? Did any prespecified data quality checks, positive controls, or tests of intervention fidelity succeed? Are any additional post hoc analyses justified, performed appropriately, and clearly distinguished from the preregistered analyses? Are the conclusions appropriately centered on the outcomes of the preregistered analyses? Are the overall conclusions based on the evidence?

Many thanks for raising these issues. We adhered to all the requirements of PCI registered reports and did not deviate from the approved protocol. We conducted ONLY the approved analysis which can be verified via the link to the approved protocol posted on the first page of this manuscript.

Overall, in my view, these questions are answered affirmatively. I just have a few small questions, largely revolving around a couple of questions from my Stage 1 review that I feel still could be addressed better.

Thanks for taking the time to review our Registered Report. We appreciate your time and effort.

-One other question first though. It seems like the authors did a great job sticking to their preregistered protocol and analysis plan. Just to make it clear for readers, though, could the authors clarify if there were any deviations from their Stage 1 proposed protocol or analyses? Again, it seems that the answer is "no", which is great, but stating that explicitly could further reassure readers about the transparency of the methods and results.

Many thanks for drawing our attention to this issue. We have now included such an explicit statement with a link to the accepted Stage 1 protocol on page 21.

-For the preliminary study, the authors said in their Stage 1 response that they reported the Nakagawa R^2, but I do not see it present. Could authors make sure it is present, and if it is not, report it? I do see it reported for the replication.

Thank you for catching this error. We have added the Nakagawa R2 on p.17.

-I feel there is still more that could be said regarding external validity, particularly in the light of the replication results. Here and in the Stage 1 response, the authors did a great job of justifying the decision to use the economic self-interest decision-making task compared to a verbal interview and included a nuanced discussion of the trade-offs between the two methods. My point is different than that, though. Instead, my point (also raised in Stage 1), is that any sort of artificial lab task will have difficulty capturing the threats (costs) caused by a real-life disclosure decision in which one's life or freedom (or the life or freedom of

loved ones) is at risk. The difference in severity of the costs of disclosure in real criminal scenarios compared to the economic costs used in the task could also partially explain why attention to benefits seemed to be the bigger driver of decisions here compared to the proposed model. As the authors mention, this sort of artificiality is common and necessary to examine the issue psychologically, so it is not a huge deal, but I still would appreciate a sentence or two to this point in the "external vs. internal validity" discussion subsection.

Thanks for insisting on this point. Upon reflection, we realize that it is necessary to comment on this limitation with costs and risks. We have now included two sentences (page 34) to make the point explicit.

"we must admit that the possibility of not winning a few hundred US dollars pales in comparison to the costs interviewees might face. This limitation with external validity is an issue investigative interviewing research needs to surmount with creative research designs."

I thank the authors for their time and efforts.

Thank you for taking the time to review our registered report. We appreciate your investment in our work.