PCI Registered Reports #315: Managing Disclosure Outcomes in Intelligence Interviews

Dear Zoltan,

Many thanks for the opportunity to revise and resubmit the captioned registered report. We are also grateful to the reviewers for investing the time and effort to assist us in improving the proposal. We have now revised the registered report according to the concerns.

Sincerely,
Authors

Recommender comments

Once again sorry for the far too long delay in getting back to you. In the end I asked 60 potential reviewers - and three responded who were experts in relevant areas, and I was very pleased those were the ones we now have. They are largely happy with your proposal, but have various comments concerning details of the protocol and requested clarifications.

We completely understand the difficulty in securing reviewers. Thanks for persisting, we appreciate your efforts. We have now revised the registered report according to the concerns. Next follows our point-by-point response to the editorial comments. To avoid confusion or misquoting, we have included the comments verbatim (in black font), and our response follows in red font.

Table of Contents

Reviewed by Jason Chin, 22 Mar 2023 05:12

Title: Managing Disclosure Outcomes in Intelligence Interviews

Recommendation: Revise and resubmit

This stage 1 manuscript will test the effect various costs and benefits (and probability of costs and benefits) on the decision to disclose information.

This study can usefully provide support for the prediction that people are sensitive to costs and benefits presented in the way the study presents them. I hope the below comments will help the authors plan and report the study in the most effective way possible.

1. Conceptually, and I'm not sure this is really that important to approving the study so that it can go forward, the biggest limitation seems to be that the context is very different than in the field. For example, I did run through the full sample provided (https://samgu.eu.qualtrics.com/jfe/form/SV_ebQV2G7I90nGwu2). I found myself largely ignoring the context (the gang story) and choosing primarily based on the percentages presented to me, to maximize my points. I might be unusual, but I'm not sure many people will get caught up in the facts of case. To be fair, the video was not in my native language, so that might have been a barrier. In any case, I suggest a stronger limitation section in point 8.

Many thanks for this comment, it is a fair one. We do acknowledge that some people might ignore the narrative content and treat the protocol purely as an exercise in maximizing points. But that possibility has not been the general case and the results of the preliminary study attests to that fact. "The addition of random effects suggested the presence of meaningful individual differences in the propensity to disclose information and the sensitivity to the costs and benefits of disclosure. Indeed, as can be seen in Table 1, the random effect variance for both intercepts and slopes is considerable. In addition to the variance associated with individual people, different items of information vary in their likelihood of being disclosed. This result suggests that the narrative content of the item had a substantial influence on disclosure.

Taking points 1 and 8 into consideration, we have now revised the Internal-External Validity section into a Constraints on Generality section (pp 27 - 28). Here, we expand on the limitations of our study in relation to the verbal and nonverbal nuances, and strategic passive disclosures. We have now clarified that our results may only speak to mechanisms underlying what interviewees ACTIVELY choose to disclose.

2. Twice during the experiment, it asked me if I would accept the free points I was given. Was this mentioned in the protocol? What's the purpose of it?

The experiment referred to here was the Preliminary Study, not the protocol under consideration. In the preliminary study, we implemented the "decision to accept free points" as a means to break the potential monotony of passively reading instructions. Upon further consideration, we realized that choice was unnecessary and it is not included in the Main Study.

3. During the study, it prompted me that I may like to imagine myself in the situation but then later use mandatory language. This should be consistent.

Duly noted with thanks. We have made the language consistent throughout, instructing participants to assume the role of the main character.

4. Page 21: The authors say that the first study suggests its sample was sufficient to detect "the effects of interest". I think more precision is needed here. What's the inferential goal? Do they mean effect *size* of interest?

Thanks for drawing our attention to that oversight. We meant to say: the effect sizes of interest. We have now edited the text accordingly and referred readers to the Analysis Plan, which contains the corresponding power analysis.

5. Page 22: The authors say they have two memory checks that they will use to exclude "inattentive participants". Does this mean that if you fail both you are out? Or just one?

Similar to the Preliminary study, we will exclude the data of participants who fail one or both memory checks. We have edited the text to eliminate ambiguity.

6. Page 25: The authors say there were no anomalies in the data. How do they operationalize that?

Apologies for the imprecise language. Here, we meant to say that we observed no floor or ceiling effects and that the majority of participants understood the instructions of the protocol. We included that sentence due to previous reviews wherein reviewers claim that participants might not understand the instructions. Upon reflection and based on the illustrations we added to this version of the manuscript, we believe that sentence is unnecessary and have deleted it.

7. Page 26: The authors explain the inferences that will be drawn from their data and analysis: "To support the hypotheses, consistent with the preliminary study, the coefficient for benefit should be positive, and the interaction should be negative." This is supplemented by Table 3 (study design template), which explains that non-significant effects cannot help disprove the disclosure model or elements of it.
    a. I would include in and around page 26 more elaboration on these points. Can the authors say anything more precisely about what statistically significant effects will tell us about the model? Would it help to list out the hypotheses more clearly?

Thanks for this comment. At this stage, we cannot elaborate on this issue beyond what we have noted on p. 27: "…the study will not have adequate power to detect effects that are substantially smaller (than those found in the preliminary study). Because of this limitation, if the results are nonsignificant, we will not be able to make claims about the absence of theoretically relevant effects. The results of the proposed will provide insight into the possibility of smaller effect sizes then we might be able to theorize about such smaller effect sizes and design further studies where we can make claims and test for the absence of effects. We are happy to make further revisions suppose we have misunderstood the point the reviewer is making here.

8. Pages 27-28: Limits to internal and external validity are described, but I think more could be said when it comes to realism (see my first point). This could be in the format of a constraints and generality statement.

Kindly see response to Comment 1.

9. I assume the planned study will be registered, with its data, code, and materials made openly available?

Absolutely, we will make everything available in line with the requirements of PCI RR.

I always sign my reviews,

Jason M. Chin (ORCID: 0000-0002-6573-2670)

Thank you for taking the time to give us comments, we appreciate your feedback.

Reviewed by Yikang Zhang, 06 Mar 2023 20:42

Managing Disclosure Outcomes in Intelligence Interviews is a stage-1 registered report focusing on empirically testing the proposed model of disclosure decisions in intelligence interviews. In the preliminary study, the authors examined how estimated cost and benefit of disclosing specific information could impact the disclosure decisions. Consistent with the proposed model, 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. The materials, data, as well as analysis scripts are available for inspection and consistent with the descriptions in the manuscript. The proposed study is a conceptual replication of the preliminary study with several featured changed. The materials as well as analysis plan are shared and reported.

Overall, my opinion is that the manuscript is well-written and the potential findings are of scientific value. Therefore, I believe that this work has the potential of being recommended by PCI RR. However, I did notice that there are certain minor issues that could be further improved. Therefore, I suggest a minor revision for the current submission.

Thank you for taking time out of you schedule to give us feedback. We appreciate your comments.

Below please see my more specific comments.

1. The authors mentioned individual differences in disclosure. The package *jtools* has a function *summ* can provide the intra-class correlations, which maybe more intuitive than variances.

We appreciate this suggestion. We have added ICCs to the presentation of the preliminary study (calculated using the performance package rather than jtools, but the math is the same).

2. In addition, visualization of the random effects would be even better for reader's understanding (not a must).

We agree that a visualization of the random effects might be useful. For the preliminary study, we have decided not to include such a visualization (mainly to avoid cluttering the presentation). However, for the primary study, such a visualization may indeed be useful as an exploratory tool and a way to display the results to the reader.

3. Is there a measure to ensure participants will complete viewing the video?

Many thanks for drawing our attention to this technicality. To ensure that participants watch the entire video. We will disable the advance button for the length of the video.

4. The authors stated that bimodal distribution is most obvious in the high-stake condition, consistent with the model. I was wondering why wouldn't we see the pattern in the top panel of Figure 2. From the top panel, it seems that participants disclosed various

numbers of details ranging from 1-6 (or maybe I misread the Figure).

Clarifying this point is indeed important. We have added a footnote expanding on this specific difference in the distributions. Here is the text of the footnote: "Note that the upper panel of Figure 2 does not feature similar bimodality. The upper panel displays disclosure frequencies for each participant, across all scenarios they encountered, whereas the lower panel displays proportions of disclosure for participants decisions in each scenario. Thus, for high-stakes information, in any given scenario, participants seemed to make a decision to disclose no such information or a substantial amount of that information, producing a bimodal distribution at the scenario level. From scenario to scenario, however, a participant might have been variously highly forthcoming or highly withholding with high-stakes information, producing a different distribution at the participant level."

5. The authors mentioned generalizability as a goal of the proposed study. It does not seem obvious to me that the proposed study examines whether the effects could be generalized to different settings/contexts or different populations or methods. From my perspective, the proposed study aims to replicate the findings in the preliminary study with a highly similar method and materials while fixing certain factors that potentially could have impacted the results (e.g., the mismatch of details and cost/benefit weights).

This is a fair point. Thanks for drawing our attention to it. We agree that at this early stage our findings can only speak to replicability, not necessarily generalizability. We have edited the manuscript to reflect this state of affairs. We stick to replicability and eliminate mention of generalizability.

6. Could the authors provide the scripts for the simulation-based power analysis on OSF?

This script is now available in the OSF repository: https://osf.io/5rbu6/.

7. The texts of each scenario involved in the proposed study are quite long. Has the authors planned measures to ensure attentiveness, other than the memory questions at the very beginning? This could have an impact on the experiment results, especially when the study is a online study. Although, I can also understand that the results from the preliminary study is reassuring to a certain extent.

Currently, we believe participants stay on task and read the scenarios. But we are also sympathetic to the reviewer's point. If the editorial team agrees, as exploratory analysis, we could examine how time spent on each scenario page influences the results?

8. Is there any other compensation scheme other than the lottery?

No, the only compensation is the lottery, and we inform participants accordingly as we did in the preliminary study.

Reviewed by Tyler Jacobs, 19 Mar 2023 20:20

In this article, the authors propose and aim to test a new framework for understanding when people will disclose information during an intelligence interview called the Disclosure Outcomes Management Model. This model frames the decision to provide information as a self-interest dilemma in which the interviewee must balance the potential benefits (e.g., community safety, upholding morality) of disclosing risky information with the potential harms to the self (e.g., retaliation from the group being reported on). They then describe the calculations of risk and benefit result in four categories, with low-stakes and guarded information (theoretically) being less likely to be disclosed, unguarded information being more likely, and high-stakes information being variable. The authors then report a preliminary study that supported these hypotheses, and then begin the registered report for the planned study.

Strengths:

-The theory is described well in the Introduction and Figure 1 displayed the model in a clear way.

-The authors use sophisticated multilevel models to account for random effects in their complex experimental design, which is important for their research design.

-In my opinion, the design has strong internal validity.

Concerns and Comments:

Overall, I think that this is a reasonable design with strong proposed analyses. I do not see any major issues. However, there are few aspects that I would like the authors to consider.

-First, on the theory-side, I found the definition of "self-interest" to be unusual ("broadly encompass[ing] any outcome an interviewee may want to achieve or avoid"). In social psychology, self-interest is typically defined as the motivation to achieve outcomes that benefit the individual and avoid those that do not (Miller, 1999; Gerbasi & Prentice, 2013). Additionally, many theories specifically state that if the outcome is intended to primarily benefit others, it is not self-interested (Cropanzano et al., 2005; Holley, 1999). Thus, I would argue that "act[ing] in the best interests of other associates" is not actually self-interest. Their definition is closer to a purely economic definition of self-interest (maximizing one's gains, minimizing losses), but I am not sure that this fits either, and the interview situation is not a purely economic one. Could authors either provide further justification and citations for their definition of self-interest, or consider if another term would fit better?

We acknowledge the important point the reviewer is making here. Using common terms ensures consistency across literature; suppose the concepts and constructs are indeed identical. Our definition of self-interest might be viewed as more common in the economics than social psychology literature. Indeed, we scoured the literature for the most appropriate term and definition to describe the nature of costs and benefits according to our theorizing. And the choice is deliberate in service of avoiding ambiguity and redundancy. As noted in the manuscript people may disclose information in light of personal costs and benefits or in service of others. One might argue with the use of "self" to capture all those possibilities, but we maintain that the term is suitable because the self-interest dilemma as we define is a

higher-order objective that encapsulates motivations (i.e., sating the interviewer's information objections or refraining to do so). In the end, an interviewee has to make a *personal* decision about what to disclose even if the behavior is in service of others. Thus, we believe our definition of self-interest dilemma keeps that *personal decision* in focus while clarifying that an interviewee's behavior could be in consideration of others. Drawing on expected utility (Savage, 1954) to theorize about interviewees' decision-making is already used by investigative interviewing researchers (Yang et al., 2017), which keeps us consistent with the literature. We will be happy to raise the issue of terms again in the discussion after data collection. Then we can highlight the points the reviewer is raising here because it is worth clarifying to readers who might have the same contention.

-Could the authors more clearly state their exclusion criteria (e.g., how much missing data is too much or how many memory checks can be failed)? The authors could also consider reporting the results without exclusions in the Supplementary Materials. I would reccomend this especially given the large number of excluded participants in Study 1 (transperency would be best).

Absolutely, we agree and in line with PCI RR's guidelines all our data will be public. In fact, the preliminary study's data is already public along with analysis scripts and the excluded data (see p. 19): https://osf.io/5rbu6/?view_only=1db497ff4e7c4f6cb9d2aeb7c5b177c7

-For Study 2, given the large number of excluded participants in Study 1, could the authors describe the number of participants they will recruit (before exclusions) in order to meet the minimum sample size?

As noted on p. 22, We will include two memory checks to flag and exclude the data of inattentive participants who fail both or one of the checks. And the minimum number of participants we will accept is 300 according to our power analysis. Thus we will end data collection when we have 300 participants who pass both memory checks.

-In lieu of an a priori sample size analysis in Study 1, could authors perform a sensistivity power analysis ?

Kindly see the Analysis plan for the replication study where we conduct an a priori power analysis to ensure the replication study is well powered to detect the effects sizes that emerged in the preliminary study. In principle, it is possible to conduct a sensitivity power analysis, but this approach would still require us to make assumptions about the random effects variances. We believe the most appropriate approach is to use the results from the preliminary study to inform an a priori power analysis. We could also conduct a power analysis informed by the results from the forthcoming study, but we believe this approach has limited value.

-In addition to the model fit stats (AIC), could the authors report effect sizes for their models (R^2 or f^2; or ICC for random effects)?

For the preliminary study, we have added ICCs and Nakagawa R-squared values. We will do the same for the planned study.

-I appreciated the authors' discussion of internal and external validity. However, I would note that concerns about external (and construct) validity go beyond just how interviews would have more psychological realism. In a real-life interrogation situation, the consequences go far beyond collecting points and competing to receive a monetary reward and could involve fear for one's life, fear for loved ones, fear of implicating one's self in a crime, etc. Thus, despite the efforts to vary the consequences and incorporating a choice structure with incentives, this design likely does not perfectly capture the construct of these situations in the real world. That being said, this artificiality is common in psychology research and is necessary for internal validity. This idea, though, could be added to the discussion on lower external validity.

Many thanks for raising this issue, which was also flagged by another reviewer. And we have now revised the section on internal and external validity to a discussion on the constraints of generality (pp 27 - 28). Here, we expand on the limitations of our study in relation to the psychological realism, verbal and nonverbal nuances, and strategic passive disclosures. We have now clarified that our results may only speak to mechanisms underlying what interviewees ACTIVELY choose to disclose.

I thank the authors for their efforts, and hope that this feedback is helpful.

We appreciate the time and effort you put into reviewing our work.