

21st July 2022

PCI Registered Reports #188: How Intelligence Interviewees Mentally Identify Relevant Information

Dear Zoltan,

Many thanks for the opportunity to revise and resubmit the captioned registered report. We are also grateful to both 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. 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.

Sincerely,
Authors

Table of Contents

Editor:	1
Anonymous reviewer, 11 Apr 2022 21:21	2
Anonymous reviewer, 23 May 2022 07:33	7

Editor:

Two reviewers have seen your submission and are generally positive. They suggest some clarifications and some procedural adjustments. One issue they bring up I would like to comment on: While your N has a practical upper limit, you can still show how severely your study tests your hypotheses by simulating with your N how often the HDI would be inside, straddling, or outside the ROPE, given either that a predicted effect size is true, or that the ROPE is true (see <https://psyarxiv.com/yc7s5/>) so it is clear how adequate your N is (e.g. is it enough to squeeze an HDI into the ROPE?). As per the last reference you might also see if you can justify more precisely your minimal interesting effect size. While your suggested value seems reasonable given the effect size of the past studies you quote, one previously recommended heuristic (see ref above) is to choose the lower limit of the 95% CI of the effect in relevantly similar studies (i.e., probably the ones you cite).

Many thanks for the suggestion to use simulations to examine the level of precision our chosen sample size can provide. We have now run some simulations and raised our sample size choice to $N = 200$ participants per experiment. The simulations indicated that 200 participants

per experiment sufficed to reach the precision we desire under two conditions: (1) a scenario including a wide range of estimated effect sizes using a random draw of response probabilities; and (2) a scenario where there is no effect.

Our desired level of precision was that the width of the coefficients' 95% HDIs should be equal to or smaller than 0.5. The utterance experiment (Study 2) served as the benchmark because that study includes four outcome levels: no comment, bare details, medium details, and complete details. The designation experiment (Study 1), on the other hand, involves three outcome levels: bare details, medium details, and complete details. Hence, Study 2 versus Study 1 might require a larger sample size to achieve acceptable precision even though both studies employ a similar design. The Analysis Plan section provides a detailed description. And the code for the simulations is available here: <https://osf.io/7tgwe/>.

Suppose we were to follow the recommended heuristic in Dienes (2021). In that case, we would have chosen the lower limit of the 95% CI of effects found in similar studies, for example, Lorson, et al. (2021). Then a negligible effect would be anything less extreme than 0.1. However, given our resource constraints, we will not be able to reach a precision with HDIs equal to or smaller than 0.2 in range.

Anonymous reviewer, 11 Apr 2022 21:21

In their pre-registered report "How Intelligence Interviewees Mentally Identify Relevant Information," the authors tackle a very pertinent question within the investigative interviewing literature; namely, how interviewer questions impact interviewees' mental designation of information items.

I believe that the authors do well in discussing the importance of the topic and the related theory, as well as have developed a good procedure to test their postulations. However, there are some concerns that I have, primarily with their methodology.

We are grateful for the time and effort you spent reviewing our proposal. The concerns you raised have assisted us in sharpening our methodology. In what follows, we respond to all the comments and clarify the issues of concern.

1. Considering that the authors make hypotheses concerning resistant interviewees, I would like to see them more extensively link the counter-interrogation literature to their hypotheses (especially in regards to Study 2). For example, I would argue that consideration of theory underlying the Strategic Use of Evidence (SUE) would be important to consider (e.g., interviewees think about what information an interviewer has and uses their predictions to guide their interview strategies).

Many thanks for calling our attention to consider counter-interrogation strategies in relation to our hypotheses on resistant interviewees. As the reviewer correctly notes, resistant interviewees are likely to engage in strategic maneuvers regarding disclosure. However, we

deliberately refrained from including in-depth discussions about the strategic aspects of disclosure¹. Our choice is in service of clarity; we want to avoid confusing readers. Thus, the manuscript's introduction focuses on pragmatic considerations: the mental processes of identifying relevant information.

Nonetheless, when situating the current theory (p. 11 - 14), we mention how resistant interviewees might engage in strategic behaviors after mentally flagging relevant information. They might withhold information, disclose partial bits, or deceive. We refrain from fully delving into downstream consequences like how perceived interviewer knowledge might affect interviewees' strategizing (e.g., SUE and Scharff). We believe such a discussion will better suit the article's discussion section when we discuss the implications of the theory, given the results. **The present theory tackles what happens before strategic considerations: flagging relevant information.** (See speculative discussion of the Scharff technique at the end of this comment, for example)

We also appreciate the reviewer's comment that counter-interrogation strategies might especially feature in Study 2. The design of Study 2 deliberately keeps the possibility of complex strategies to a minimum via short narrative scenarios that limit the range of utterance choices (i.e., potential discourse moves). We do not aim to examine counter-interrogation strategies in the present work. Nonetheless, the research design still allows resistant interviewees to demonstrate resistance in their utterance choices. As we note in Hypothesis 2c, resistant interviewees should exhibit the following preference pattern: no comment > bare details > complete/medium details, regardless of question-type. Study 2 complements Study 1: our goal is to examine the contention that question-type influences mental designations (Study 1), but disposition affects utterance choices (Study 2).

Speculative discussion on strategizing

The current theory possibly sheds light on the mechanisms of the Scharff technique. The technique is an intelligence interviewing method consisting of a collection of tactics that work in tandem to facilitate interviewees' disclosure. For an overview, see Oleszkiewicz (2016; see, also, Luke, 2020). The extant evidence on the Scharff technique suggests that it promotes disclosure when an interviewee is semi-cooperative (e.g., Oleszkiewicz et al., 2014)—that is, an interviewee motivated to both disclose and withhold information.

Compared to asking direct questions, the Scharff method's solicitation attempts purposely obscures an interviewer's information objectives (Granhag et al., 2019; Oleszkiewicz et al., 2014). Based on the theory this article presents, one can speculate that the Scharff method leads a semi-cooperative interviewee to mentally designate a broader range of potentially applicable information items. The interviewee wants to disclose at least some information, but she does not know what the interviewer is actually after. This effect results in the mental designation of a larger pool of items the interviewee could possibly disclose. Conversely, direct questions may produce a narrower range of applicable information items. Such questions better

¹ In previous reviews, reviewers have erroneously assumed the article was offering a theory about the strategic aspects of disclosure and non-disclosure.

allow the interviewee to identify the interviewer's objectives. Consequently, if an interviewee is semi-cooperative, and the Scharff technique (vs. direct questioning) produces a broader range of possible information items to disclose—then the Scharff method is likelier to influence interviewees to disclose more information.

We are not asserting that the Scharff technique derives its efficacy solely from obscuring interviewer objectives. Indeed, the method consists of various components that function together to facilitate disclosure. Our contention is that the way the Scharff technique frames questions leads interviewees to mentally designate a larger pool of information items. This effect allows the method, compared to direct questions, to inadvertently solicit a wider range of information items from semi-cooperative interviewees.

2. The authors should consider including a sensitivity analysis for their restricted N of 300.

Thanks for this suggestion. We have now conducted such a sensitivity analysis via simulations as the editor suggested. The simulation led us to raise our sample size to 400, 200 per experiment. Please see our response to the editor for a detailed description.

3. Concerning their design (p. 17), the authors should clarify that it is a between-subjects design. Additionally, the authors should reiterate what the primary outcome variables are (which highlights the key distinction between Studies 1 & 2).

We have now clarified the design more explicitly: "The experiments will employ the same design: 2 (Question-type: low- vs. high-worthwhileness; within-subjects) × 2 (Disposition: cooperative vs. resistant; between-subjects)."

On page 16, we mention the primary outcome variables in our studies. "Study 1 (N = 150) will focus on designation choices and Study 2 (N = 150) on utterance choices."

4. What is the authors' reasoning for presenting information items individually instead as part of a narrative (which would, arguably, increase the ecological validity of the study)? By presenting items individually, the authors are presumably making an initial decision for participants concerning which information out of a narrative to hone in on (and this is especially so for the low-worthwhileness questions), thereby undermining, to some extent, the premise of their study.

Because the present work is the first test of the theory under contention, we designed the studies to maximize internal validity. However, the design does not eliminate ecological validity, given our aims. We deliberately employ concise narrative scenarios that allow for a limited set of discourse moves. Those discourse moves are represented in the multiple-choice options. As noted on page 20, interlocutors typically restrict their discourse moves to the issue under discussion (see, e.g., Roberts, 2012). Granted, the multiple-choice options are complete sentences, and people do not necessarily converse with complete sentences. However, the multiple-choice options embody the essence of any discourse move one can make in the

scenarios—given the interviewer questions. Hence, it is unlikely that participants will hone in on or generate information items other than those in the multiple-choice options. The multiple-choice options do not make an initial decision for participants. By outlining all the possible discourse moves, the multiple-choice options offer two design advantages. (1) They speed up the time participants would spend on the experiments. We believe this approach is ethical and prudent, given the compensation we can offer participants. (2) Also, the multiple-choice options eliminate potential flexibility and errors that may arise when coding free text responses.

The above comment notwithstanding, we appreciate the reviewer's concern. The present work is a stepping stone. In future studies, we intend to further enhance ecological validity by implementing complex narratives and allowing participants to generate information items from scratch.

5. What is the authors' reason for including the wager question? Also, is this a continuous variable, or dichotomous variable?

The wager question will serve as an auxiliary confidence measure to assess the reliability of the direct confidence rating. People place higher wagers the more confidence they have in their predictions.

Apologies for failing to clarify that both measures will be examined on a continuous scale; we have corrected that issue (p. 20).

6. Have the authors considered enhancing participants' motivation? For example, the authors could tell participants that their final payout is dependent on: how well they predict what the interviewer is looking for (Study 1); or how effective they are at demonstrating their innocence (Study 2)?

We considered adding extra motivations to the disposition manipulation but decided against it for now. As discussed previously, the present work focuses on the first aspect of navigating an interview: the mental processes of flagging relevant information. Our focus is not the second aspect of navigating an interview: strategic disclosure. We suspect that extra motivations might induce such strategizing to confound our findings. Additionally, we want to keep the design of Studies 1 and 2 as identical as possible so we can be confident that question-type and disposition induced any difference in effects, not extra motivation.

The decision-making IMCs and control questions will assist us in detecting lackluster responses.

7. What is the reason for the authors allowing participants to miss 2 (instead of only 1) of the control questions? These seem like extremely straight-forward questions, making me wonder whether an allowance of 2 misses is too lenient.

Many thanks for flagging this possibility. We agree and will now exclude those who fail one control question (p. 21).

8. This is related to my comment above about the wager variable – why are the authors running a logistic regression? I assumed the variable was continuous (0-100%). If it's not, the authors should clarify this.

Confusion may have occurred here due to the issue the reviewer pointed out in Comment-5, and we apologize. As noted on page 20, the wager question will be optional; this feature will elicit two variables. (1) The number of participants who place bets versus those who do not bet. (2) The wagers of those who place bets. Thus, we can examine item-1 “willingness to bet” using logistic regression models, (noted on page 28). And we can explore how the relationship between item-2 and the mandatory confidence measure.

9. I am curious about the authors' decision to frame the responses to the disposition manipulation check as they did (p. 1 of the Appendices). I can see a resistant subject answering with (-1), (0), or (1), which I do not believe that any of those answers would indicate that the participant did not understand the instructions. For example, the resistant subject may be motivated to be viewed as cooperative, and therefore adopt the strategy of revealing some information, or use a mix of commission and omission strategies.

Furthermore, I can see a cooperative subject answering with (1) or (2), as they may be fearful of something like gang retaliation, which may therefore lend to a motivation to resist.

Have the authors considered these possibilities? I suggest that the authors include a more concrete manipulation check to base exclusions on.

These are excellent points, and we have considered the possibilities the reviewer raises. There is a difference between an immutably resistant interviewee and a semi-cooperative interviewee who is willing to reveal some information. Thus, as the reviewer points out, resistant interviewees may vary in their resistance level. Additionally, cooperative interviewees may differ in their level of cooperativeness. But it is worth noting that the measure in contention is a stimulus manipulation check, **not an instructional manipulation check (IMC)² by which we will exclude participants**. We will examine the level of cooperativeness on a 4-point bipolar continuous scale ranging from lying (-1) to full-scale cooperativeness (2). **The aim is to check the efficacy of the disposition manipulation by examining the mean difference between the cooperative condition and the resistant condition.** We expect the cooperative condition to elicit a significantly higher level of cooperativeness.

Indeed, creating a manipulation that perfectly instantiates immutable resistance and full-scale cooperativeness is challenging. The level of cooperativeness is bound to vary. The “best” manipulation would flat out tell participants to be immutably resistant or fully cooperative. Such a manipulation will create demand characteristics and confound our results. Participants will simply do what we ask them to do. In the real world, people decide on their level of

² IMCs are purposely designed to flag, for example, lackluster responding and exclude such participants.

cooperativeness based on the circumstances in which they find themselves. Thus, our stimulus will create the situation, and we will measure participants' dispositions as a function of the stimulus. That approach will help us assess the robustness of our manipulation.

10. I would advise against including compound questions (e.g., Question #7), as such questions will presumably require subjects making multiple information designations instead of just one.

11. Some of the low-worthwhileness questions seem overly vague in relation to the piece of information they are paired with (e.g., Question #1), as compared to others (e.g., Question #2). I suggest that the authors consider more specific low-worthwhileness questions across the board (e.g., as they used with Question #2).

To elaborate, the low-worthwhileness question for #1 would give very little (if any) indication that the interviewer is looking for information about the type of drug the gang is selling. However, with Question #2, the interviewer is arguably giving more indication what they are looking for (i.e., something about the drug deals specifically), while still being vague enough to be in line with the manipulation. Thus, as it stands, the extent to which a question demonstrates low-worthwhileness is confounded with the piece of information.

For conciseness, we tackle points-10 and -11 together. Thanks for bringing these issues to our attention; the remarks are very much appreciated. The concerns are legitimate, and we have implemented revisions to address them. We have edited all compound questions out. Additionally, we have ensured that all the low-worthwhileness questions relate to the specific scenarios (see Appendix C).

Anonymous reviewer, 23 May 2022 07:33

This is a very thorough consideration of the planned research, and the associated materials and analyses. The research questions make sense considering the theories that are described, and the hypotheses are clearly stated and defined. Based on the study design and analyses, participant responses should provide data which will allow the research question and hypotheses to be answered. The procedural details, and the materials in the appendix, are especially detailed, and will easily allow replication by an expert. This assumes that an example question is stated. The analysis is stated step-by-step with no obvious gaps (or flexibility) in relation to the tests that should be run. The only caveat is that whilst I am familiar with most of the proposed analyses, there are elements that I am not familiar enough with to comment (Priors and ROPE). There is clear evidence of what evidence will, or will not, support the hypotheses. There are manipulation checks incorporated in the study design, and attention checks to identify participants who are not fully attended to the task. Finally the report offers a novel study, with an interesting approach that I am looking forward to reading once the research is published.

There are just a couple of details that I would like to raise:

Thank you for taking the time to review our work, for the kind words, and for calling our attention to essential details. Next, we respond to the comments and clarify any outstanding issues.

(1) The report states that resource constraints dictated the sample size. Whilst I completely understand this, it might be helpful to indicate how far away from an ideal sample size the proposed 150 (x2) sample is. If the difference is minimal, then this can be mentioned as a rationale. If there is a big difference then this really needs to be justified

Thanks for this suggestion. We have now conducted such a sensitivity analysis via simulations as the editor suggested. The simulation led us to raise our sample size to 400, 200 per experiment. Please see our response to the editor for a detailed description.

(2) I am slightly unsure about the claim that there should not be a response preference in the Cooperative/low-worthwhileness conditions.e.g. hypothesis 1b and 2b. Having read the low-worthwhile questions for the first time, and assuming that I am a cooperative interviewee, I would have tended to choose the complete response option each time. This could be my own personal disposition, but this might be worth thinking about when interpreting the data.

The reviewer's point here is indeed an issue to keep an eye on when interpreting the data. At this time, our hypotheses are predictions to be examined. The reviewer's speculation might well be the case. But we can only decide on the issue with more certainty after data collection and revise future hypotheses accordingly. As we noted in our responses to Reviewer-1, levels of cooperativeness between participants might vary. Thus, we structured the current predictions about low-worthwhileness questions as we did. Given that the (1) low-worthwhileness questions do not specify an objective, and (2) assuming that cooperativeness levels between interviewees might vary, (3) we propose that different interviewees might think the interviewer wants to know different aspects of the discovery. (4) Thus, our prediction that low-worthwhileness questions will not elicit particular response preferences. This issue is one of the reasons we include the confidence ratings to assist us in interpreting the findings.

The reviewer's comment also highlighted the prudence in revising option-4 of the disposition manipulation check. Previously, option-4 was "I will reveal *everything* I know". We have now changed option-4 to "I will reveal *what* I know". We believe the change provides a more neutral assessment of cooperativeness without influencing participants to flag complete details in the discoveries.

To reiterate, we do not take the reviewer's comment for granted. We will be vigilant about the concern when interpreting the findings.

(3) There are a couple of language/spelling errors in the materials that will need to be addressed .e.g. under confidence measures via bets (Appendix, pg.6).

Many thanks for calling our attention to errors in the Appendix, apologies for the oversight. We have now proofed the Appendix and corrected all errors.

(4) Apologies if I missed this, but I presume that when asking participants about what % of their compensation they are willing to bet, that this is not an actual bet. How is this explained to participants, and how might this effect their responses. If this is an actual bet, then this has potential ethical implications.

Yes, the willingness to bet measure is hypothetical. We have now edited the instructions in the Appendix by explicitly informing participants that the wager is hypothetical. The bet is an auxiliary confidence measure to assist us in assessing the reliability of the direct confidence rating.