Dear Dr. Zhang,

Thanks for inviting us to revise our submission. We have marked all our edits in this submission using TRACK CHANGES.

Sincerely,

Authors.

Table of Contents

Editor

Dear authors,

Thank you for submitting your Stage 1 registered report '*How Interviewees Determine What Interviewers Want to Know*' to PCI RR and for your patience during the reviewing process. I have now received three independent reviews from experts in this field. Based on these reviews and my own reading of your manuscript, I would like to invite you to revise the proposal. There is much to like about the manuscript already (e.g., the sharing of materials), but I will highlight the most salient opportunities for further improvement:

As noted in the cover letter that accompanied our initial submission, we collected the data for Replications 1 and 2 in June 2024 due to administrative constraints. But we have not analyzed the data in any way. We will conduct the analysis after a decision has been made.

We appreciate the reviewers' time and effort in engaging with our proposal. In what follows, we provide point-by-point responses 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.

To facilitate a snapshot of our responses to the reviewers' comments, we will repeat some relevant responses to the three points below.

1.   As pointed out by the reviewers, in the introduction, much relevant literature from the field of investigative interviewing and memory research are missing. One reviewer also pointed out that the assumption in the current model that the interviewees assume that they need to tell the complete information is at odds with the assumption from Cognitive interview.

Relevant literature and theoretical basis: Indeed, we have written extensively about the theory's rationale elsewhere (i.e., Neequaye & Lorson, 2023) and wanted to avoid repetition. On page 3 (last paragraph) to page 4, we have now included a summary of the Pragmatics literature that guided our thinking, namely Gricean norms and Relevance theory.

We specifically avoid discussing issues about memory to prevent confusion. Memory is relevant to the present research but distinct from the current investigation. As noted in Neequaye and Lorson (2023, p.5 - 6), "memory addresses disclosure and nondisclosure at the macro-level. That is, what makes an interviewee report accurate details on a topic, assuming the interviewee is willing to report on the topic. The present theory tackles the micro-level what happens **before**, for example, memory devises or strategic question-types might improve the accuracy of disclosure. The present theory explains how interviewees hear from a question the thing that is being questioned. Hearing what is being questioned from a question is a pragmatic matter; here, the interviewee's problem is not yet an issue of memory and recall. Our contention is that hearing what a question is questioning is an issue of pragmatics.

Compatibility with the Cognitive Interview:

Reviewer 1 comment: In fact, the Cognitive Interview is trying to counter-act the tendency of witnesses to assume that the police does not want to know all the witness may know.

Yes, the Cognitive Interview tries to elicit exhaustive detail, but nothing we have said in the current manuscript is at odds with the ethos of the Cognitive Interview. Our theory simply states that the way a question is framed can lead an interview to focus on other details to the exclusion of other information—not necessarily that question-types lead interviewees to think interviewers do not want to know some information. **Again, never in the manuscript do we claim that something is accurate or inaccurate about the Cognitive Interview**.

Importantly, we are not advocating for using or avoiding any interviewing technique. Our goal is to examine the mechanisms underlying the pragmatics of identifying what interviewers want to know.

2.   Several design choices are not clearly explained in the current manuscript, for example, the hypothesis regarding disposition manipulation. Reviewers also raised doubt about the validity of this manipulation. One reviewer also pointed out some potential issues in the high vs. low specificity manipulation, which is central to the current research program.

We have now clarified all issues with our design choices in our point-by-point responses.

Reviewer 2 comment: On this note, could there be inclusion of ratings (or ranking of most relevant disposition) in the current research to see to what extent participants accurately perceive their instructed disposition? My concern is that the semi-cooperative and resistant dispositions could be perceived similarly by interviewees as in the latter, they still need to avoid suspicion, which can induce a semi-cooperative mindset.

We believe the response options where participants rate how they should behave given their dispositions examines how accurately they perceive their instructed dispositions. For example, if participants assigned to the semi-cooperative and resistant dispositions consistently choose "I will lie to hide what I know" then it will be evident that the roles are perceived similarly. That is to say, we have an objective test of how participants perceived their assigned dispositions and whether there are significant differences between those perceptions.

Reviewer 3 comment: How the manipulation check regarding disposition would be scored wasn't stated, and I did not find it to be obvious. Further, it wasn't clear if failure to answer this "correctly" would result in exclusion or not. Presumably not, as other exclusion criteria were noted.

The scoring of the disposition check is stated in Appendix A (range -1 to +2). This variable will not be used in any exclusions, which is why we did not include it in the exclusion markers. The variable will be used to explore whether interviewees generally understand the implications of the disposition manipulation.

3.   Regarding hypothesis formulation and hypothesis testing. I agree with the reviewers that H1 could be improved in its formulation to increase readability. Second, I think it would greatly benefit the readers if the authors can explain the analytic approach as well as the simulated power analysis in a more layperson-friendly way. I appreciate that the authors shared the R Code during the review. However, I assume that it is quite effortful for readers to make sense of the analysis solely based on the code, even if they are familiar with R. Please note that if there are changes in the hypotheses and tests (be it manipulation checks or primary hypothesis testing), these should also be reflected in the Study design table.

Clarifying Core Hypothesis 1: We have now expanded on the hypothesis to clarify the meaning of pragmatic correspondence (see page 9). "When interviewees determine what an interviewer wants to know, high-versus low-specificity questions should elicit a greater preference for pragmatic correspondence (Core Hypothesis 1). That is, high- versus low-specificity questions should lead interviewees to focus more on information that pragmatically answers or corresponds to the interviewer's objective."

Analytic approach: We respectfully disagree with Reviewer 1 that our analysis strategy is overly complex and a simpler analysis would suit. Note that the reviewer did not provide any example of such a simpler analysis or an explanation to support the critique that our analysis strategy is overly complex. Unlike t-tests and chi-square tests, our strategy can handle multiple predictors simultaneously, analyze complex relationships such as interactions, control for confounding variables, and provide a more comprehensive understanding of the relationships between variables.   We followed PCI's recommendations on what an Analysis plan Section should contain and provided exhaustive detail to allow reproducibility and replication. Kindly see a similar analysis plan recommended by PCI: https://rr.peercommunityin.org/articles/rec?id=188. And we have been able to communicate the findings of that analysis strategy to a general audience (Neequaye & Lorson, 2023). Respectfully, the Analysis Strategy section is not for the lay reader but for an expert to assess.

Nonetheless, we have added a paragraph (p. 16) explaining, in lay language, why we chose the current analysis plan.

"We use linear models because, unlike t-tests and chi-square tests, they can handle multiple predictors simultaneously, analyze complex relationships such as interactions, control for confounding variables, and provide a more comprehensive understanding of the relationships between variables. Furthermore, unlike more complicated machine learning models, linear models are more straightforward to implement, explain, and validate, offering a balance between flexibility and ease of interpretation while still handling continuous and categorical data efficiently. Finally, a Bayesian approach enhances linear models by incorporating prior information, providing full probability distributions for model parameters thereby offering a more comprehensive measure of uncertainty compared to single point estimates. The approach

is robust when considering smaller sample sizes and complex models, making it ideal for nuanced real-world scenarios with hierarchical structures or varying uncertainties."

All in all, I find the insights from the reviewers helpful and I hope that the authors will engage with their comments carefully and thoroughly. The authors are not expected to accept every suggestion from the reviewers if they believe otherwise and can justify their decisions.

Many thanks again for editing our proposal.

Best regards,

Yikang Zhang

Review by anonymous reviewer 1, 29 Jun 2024 04:36

Thank you for the opportunity to review this plan for a registered study. While I found the overall research question (how interviewees determine what they think interviewers want to know) interesting and worthwhile, my enthusiasm was dampened by several issues with the proposed registered report. These were:

Thank you for taking the time to review our proposal.

-The authors would like to test whether their pilot study had a confound by replicating that within-subjects design study and then running a study with a between-subjects design and compare the results (If I understood the aim correctly). Why not just vary the presence or absence of different question types experimentally in one experiment? That would seem a much more straightforward and effective way of addressing the issue of whether the assumption that interviewers want to know all dependend on the within-subject nature of the design. Also, it seems to me that the whole setup with core hypotheses and revision hypotheses is not appropriate for a registered report that should, in my opinion, more clearly commit to a particular hypothesis/es that flow from a theoretical framework.

Thank you for this comment, but we respectfully disagree with the characterization of the research design and the implications. First, the research is not simply about replicating a pilot study - Neequaye and Lorson (2023) was a Registered Report, which is now published. Second, and more importantly, varying the presence versus the absence of question-type (i.e., low- and high-specificity questions) in the same study alone will not be rigorous. We believe that a close replication of Neequaye and Lorson (2023, Study 1) and a between-subjects study with the appropriate statistical power will better reveal whether a difference in research design is influential. The reviewer's suggestion here will only lead to speculation because one can always claim that statistical power could be an issue (note the difference in the effect sizes we used to generate sample sizes in Neequaye and Lorson (2023) versus the current proposal).

We also disagree with the comment that it is better to commit to one hypothesis over examining competing hypotheses. Philosophers of science (see, e.g., Meehl, 1990b, 1990a) recommend competing hypotheses versus testing against the null hypothesis (which often leads to speculation and post-hoc reasoning). Specifying a priori competing hypotheses with a well-powered study offers a more severe

test of claims and will be more informative when interpreting null findings — because, in our case, such a finding will support the Revision Hypothesis.

Furthermore, we show commitment to falsifying our theory by specifying Core and Revision Hypotheses. We provide a brightline for any other analyst aiming to falsify our theory. Everyone knows the conditions that will require a revision of the theory the manuscript examines.

References

Meehl, P. E. (1990a). Appraising and Amending Theories: The Strategy of Lakatosian Defense and Two Principles that Warrant It. *Psychological Inquiry*, *1*(2), 108–141. https://doi.org/10.1207/s15327965pli0102_1

Meehl, P. E. (1990b). Why Summaries of Research on Psychological Theories are Often Uninterpretable. *Psychological Reports*, *66*(1), 195–244. https://doi.org/10.2466/pr0.1990.66.1.195

Neequaye, D. A., & Lorson, A. (2023). How intelligence interviewees mentally identify relevant information. Royal Society Open Science, 10(8), 230986. https://doi.org/10.1098/rsos.230986

-I feel that the planned analyses are overly complex for the relatively straightforward and simple experimental designs proposed. Why not use more classical statistical tests? At a minimum the proposed analytical approach needs to be motivated much more clearly. Admittedly, I am not competent to assess all the choices made but then many readers will probably feel the same leading to difficulties in communicating the findings clearly.

We respectfully disagree that our analysis strategy is overly complex and a simpler analysis would suit. Unlike t-tests and chi-square tests, our strategy can handle multiple predictors simultaneously, analyze complex relationships such as interactions, control for confounding variables, and provide a more comprehensive understanding of the relationships between variables. Moreover, we followed PCI's recommendations on what an Analysis plan Section should contain and provided exhaustive detail to allow reproducibility and replication. Kindly see a similar analysis plan recommended by PCI: https://rr.peercommunityin.org/articles/rec?id=188. And we have been able to communicate the findings of that analysis strategy to a general audience (Neequaye & Lorson, 2023). Respectfully, the Analysis Strategy section is not for the lay reader but for an expert to assess.

Nonetheless, we have added a paragraph (p. 16) explaining, in lay language, why we chose the current analysis plan.

-Most importantly, I feel that the theoretical model is not explained clearly enough.

The concerns that follow this opening comment need targeted responses, which is why we have divided the paragraph into bits. But we can say something about explaining the theoretical model already here. First, a published paper explains the model exhaustively (Neequaye & Lorson, 2023). We have cited that article multiple times and provided the most relevant summary (pages 3 [last paragraph] to 5). Additionally, Figure 1 illustrates the theory even further.

The authors suggest that "Then, researchers can better specify how various interviewing methods might exert their effects." referring to the potential of understanding what the interviewer wants to know. This claim should be further explained using examples.

Given the point above, it is not entirely evident what remains unclear about the theory. The point of the sentence the reviewer has flagged is to highlight the need to understand how interviewees decide what interviewers want to know. And we provide illustrations mentioning that such insight will enhance the specification of techniques like the Scharff technique and the Cognitive Interview.

Also, the authors state "Before an interviewer poses any question, it is reasonable for interviewees to assume that the elicitation of complete details is the de facto purpose of an interview." I do not think this is at all self-evident. It would depend on the context.

This is a fair point, which is precisely why we have planned to test competing Core and Revision Hypotheses to better answer the issue rather than speculating. Testing against an unspecified null hypothesis, as the reviewer previously suggested, will only lead to more speculation—because an unspecified null is inherently ambiguous. Supporting the Core Hypotheses would indicate that context (e.g., question-specificity) matters, while supporting the revision hypothesis would indicate that context has little influence.

In fact, the Cognitive Interview is trying to counter-act the tendency of witnesses to assume that the police does not want to know all the witness may know.

Yes, the Cognitive Interview tries to elicit exhaustive detail, but nothing we have said in the current manuscript is at odds with the ethos of the Cognitive Interview. Our theory simply states that the way a question is framed can lead an interview to focus on other details to the exclusion of other information—not necessarily that question-types lead interviewees to think interviewers do not want to know some information. **Again, never in the manuscript do we claim that something is accurate or inaccurate about the Cognitive Interview**.

Finally, I did not understand core hypothesis 1: "When interviewees determine what an interviewer wants to know, high-versus low- specificity questions should elicit a greater preference for pragmatic correspondence (Core Hypothesis 1)." What do you mean by preference here?

Thanks for highlighting the potential for confusion regarding Core Hypothesis 1. We have now added a sentence to the hypothesis on page 9 to clarify (see bold text). "When interviewees determine what an interviewer wants to know, high-versus low-specificity questions should elicit a greater preference for pragmatic correspondence (Core Hypothesis 1). **That is, high- versus low-specificity questions should lead interviewees to focus more on information that pragmatically answers or corresponds to the interviewer's objective.**"

-In terms of measurement of the main variable (i.e. what the interviewees think the interviewer wants to know), why not ask the interviewees to indicate for each item in the scenario they have been given if they think the interviewer's question indicates them wanting to know this item or not?

This suggestion is precisely what we did in Neequaye and Lorson (2023). We mention it on page 6: "Participants indicated what they thought their interviewer wanted to know via a **predefined response list with three options: their discovery's bare minimum, medium, or complete details.** Furthermore, we write: "The mixture of high- and low-specificity questions plus the use of a predefined response list might have introduced a potential confound." That is, NL's design could have made the potential assumption that the interviewer only wanted to know complete details an even more salient by presenting participants with a predefined list including complete details as a choice option.

The purpose of the new design is to eliminate the confound highlighted here.

-Finally, it would seem also interesting to include the perspective of the interviewer in this line of research. This would allow investigating whether there is a correspondence between what the interviewer wants to know and what the interviewee thinks the interviewer wants to know. However, I understand that both perspectives may not be possible to include in the same (series of) experiment(s).

Thanks for this comment. We agree that examining interviewer perspectives will be interesting. This aspect of interviewing is an issue we hope to take up in the future when we have gotten a better understanding of the current theorizing—Then we can potentially develop it further.

Review by Feni Kontogianni, 23 Jul 2024 11:16

The current submission describes three replications of Neequaye and Lorson (2023) to examine which of two mechanisms under consideration better captures how interviewees (as informants) perceive what interviewers want to know based on question formulation. In Replication 1, high- and low-specificity questions will be tested in a within-subjects design, vs in a between-subjects design in Replication 2. Based on results, a third replication will use a between-subjects design and new stimulus materials to address any limitations. There is thorough consideration of the method across replications to clarify previous findings and address hypotheses; a carefully planned power and statistical analysis.The investigation of the two competing mechanisms is interesting and I can see the practical implications for the field (although contextual differences are important to consider). Some thoughts below about the link to the 'disposition' factor and questions about aspects of procedure/coding.

1.   The introduction could draw from more theoretical sources and previous research to strengthen the rationale for the research question and its operationalisation. For instance, I would expect to see reference to Grice's maxims of conversation, links to related terminology and theory in applied cognition such as diagnosticity of cues when prompting relevant information from memory, fine vs coarse reporting etc. If relevant theory has been used in previous publications as cited in the current submission, a brief summary would still be helpful to the reader.

Thanks for highlighting the need to strengthen our proposal's rationale. Indeed, we have written extensively about the theory's rationale elsewhere (i.e., Neequaye & Lorson, 2023) and wanted to avoid repetition. On page 3 (last paragraph) to page 4, we have now included a summary of that Pragmatics literature that guided our thinking, namely Gricean norms and Relevance theory.

We specifically avoid discussing issues about memory to prevent confusion. Memory is relevant to the present research but distinct from the current investigation. As noted in Neequaye and Lorson (2023, p.5 - 6), "memory addresses disclosure and nondisclosure at the macro-level. That is, what makes an interviewee report accurate details on a topic, assuming the interviewee is willing to report on the topic. The present theory tackles the micro-level what happens **before**, for example, memory devises or strategic question-types might improve the accuracy of disclosure. The present theory explains how interviewees hear from a question the thing that is being questioned. Hearing what is being questioned from a question is a pragmatic matter; here, the interviewee's problem is not yet an issue of memory and recall. Our contention is that hearing what a question is questioning is an issue of pragmatics.

2.   The rationale to examine the link between the disposition (which is instructed to participants) and the participants' reported understanding of the interviewer's goals (high/low specificity questions) is not entirely clear based on the introduction. The authors state that 'this honing process can affect whether an interviewee cooperates or resists'. A few thoughts on this point:

Thanks for raising this issue. We must reiterate that our specific prediction on disposition (Core Hypothesis 3) is that disposition should **not** influence the process by which interviewees determine what interviewers want to know. When we say the "honing process can affect whether an interviewee cooperates or resists". We mean interviewees could decide to disclose or withhold the information <u>after</u> they flag they think the interviewer wants to know. That is, (1) **question-type influences the identification of relevant information,** but (2) **disposition affects whether a person discloses or withholds information B**. Neequaye & Lorson (2023) supported this distinction.

In other words, interviewees first flag what they think is relevant to an interviewer and subsequently choose what to disclose. So, it is possible that an interviewee can <u>unwittingly</u> disclose something the interviewer was not necessarily asking for, if the interviewee mistakenly thinks the interviewer wanted such information. This is precisely what the Scharff technique does. By asking ambiguous questions, it has been found that the technique leads interviewees to disclose things they did not plan to say. This possibility is a potential ethical issue we plan to discuss depending on the findings as noted on page 5. Our plan is to replicate Neequaye & Lorson's (2023) finding to gain more confidence in our theorizing.

Nonetheless, we see the possibility of confusion and have revised the sentence to reflect our theorizing better. The honing process can indirectly affect disclosure or nondisclosure.

"how interviewers formulate questions influences what interviewees hone in on as the subject(s) of interest: this honing process can **indirectly** affect whether an interviewee cooperates or resists—wittingly or unwittingly."

Given our explanations above, we will tackle the remaining related concerns briefly to be concise.

-       based on the examples provided in the method and the particular scenario, the purpose remains that the investigator wants information that can help investigate a criminal gang (overall purpose of questioning), so isn't it possible for interviewees to assume that complete details are the most informative and the objective of the investigator's questions? My point is that the scenario and related instructions can also affect participants' understanding of the goal of the questioning so instructions could be amended and/or this might need to be the focus of replication 3 (as the issue of the stimulus material is mentioned).

Absolutely, we agree. The entire point of the research is to determine whether interviewees always think complete information is the goal or whether question-types make them focus on specific things. In Replication 3, we plan to include different themes of information in the same scenario to see if interviewees still focus on complete information when an interview requests only specific parts.

-       In real practice, cooperative interviewees may understand that the interviewer wants highly specific/complete details but may be unable to report these from memory. Some consideration to the implications of the current assumption risking cooperative interviewees being perceived as uncooperative is warranted.

As noted earlier, this research is not about memory or even disclosure, at least not directly. The research is about pragmatic considerations in interviews. So, yes, we agree that limited memory might prevent interviewees from disclosing information they would have otherwise shared. However, one can have a bad memory and still be able to identify that an interviewer wants to know about the *design* of a bomb, for example. Here is the distinction between pragmatics and memory. An interviewee could accurately flag that the interviewer wants to know about a bomb design (pragmatic issue), but the interviewee may fail to remember the design because they did not encode that information when they previously came into contact with the bomb (memory issue).

-        Participants are actually instructed from the start (or prior to questioning) to be cooperative, semi-cooperative, or resistant (which would indeed be the case). Related to this, I understand that Hypothesis 3 is essentially a null hypothesis to show that there should not be an effect of disposition on the interviewee's understanding of the question's purpose, but this seems to contradict the point above and the following one: 'Interviewees must identify their interviewer's objectives before they can determine the extent to which they might cooperate with or resist their interviewer's requests' – a decision which to be fair can be made based on multiple other factors, also depending on context.

Thanks for raising this point, but we believe the is no contradiction, and we apologize for the potential confusion. As we noted above, the present research does not examine disclosure or nondisclosure. Our focus is pragmatic considerations. We simply claim that one must understand a question's purpose **before** disposition or other contextual factors most proximate to disclosure can take effect. Identifying the question's purpose has an indirect influence because it tells the cooperator or resistor **how best** to achieve their goals. Deciding what to do, in this case disclosure or nondisclosure, is when disposition kicks in.

3.   Thank you to the authors for making the protocol available on Qualtrics. About 'S1 Manipulation check': although a manipulation check to ensure that participants understand their role based on the instructions is necessary, this check asks about their preferred strategy, which is not exactly the same. It would be interesting to collect data on participants' strategies at the end of the study, but I am not sure it can serve as a manipulation check on its own.

Thanks for raising this important point. As we have already collected the data for Replications 1 and 2 (which we noted in the cover letter [we have not looked at or analyzed the data]), we will devise an additional manipulation check for Replication 3 for review.

4.   On this note, could there be inclusion of ratings (or ranking of most relevant disposition) in the current research to see to what extent participants accurately perceive their instructed disposition? My concern is that the semi-cooperative and resistant dispositions could be perceived similarly by interviewees as in the latter, they still need to avoid suspicion, which can induce a semi-cooperative mindset.

We believe the response options where participants rate how they should behave given their dispositions examines how accurately they perceive their instructed dispositions. For example, if participants assigned to the semi-cooperative and resistant dispositions consistently choose "I will lie to hide what I know" then it will be evident that the roles are perceived similarly. That is to say, we have an objective test of how participants perceived their assigned dispositions and whether there are significant differences between those perceptions.

5.   The outcome for confirmation of hypothesis 1 could be stated more explicitly to how participants can respond in the experiment – e.g., does 'elicit a greater preference for pragmatic correspondence' mean that participants will be more likely to identify that the interviewer aims to focus on highly specific details – and how?

Apologies if this part of the manuscript was missed. Yes, we have stated how the hypotheses will be analyzed in explicit and exhaustive detail, and even how the test parameters should manifest. Please see the Study Design Template.

6.   Questions 5 and 6 (p. 34) – high vs low specific questions are formulated quite differently relative to others (with respect to the target of the question). For example, question 6b 'Have you made any observations about KET22's customers?' could refer to a personal description of the customers, vs how they contact KET22, which is a lot more specific in 6a. I understand that is part of the point, but my concern is that questions 5 and 6 are much broader than in all other question sets, where e.g., instead of 'exact location' for high-specificity, 'anything about where' is used for low-specificity, thus still obviously capturing the 'exact location'.

This is a fair point, thanks for raising it. In fact, this issue is one of the reasons why we opted to conduct Replication 3 after analyzing the results of Replications 1 and 2. Then we can use the most suited stimulus material. So, if scenarios 5 and 6 (or any other scenarios for that matter) bring too much noise to the data, then we will revise them for the review of Replication 3.

7.   Coding strategy:

-       Shouldn't there be a discussion of a limitation of the coding strategy if the ICC value is below 0.6?

We have not coded or analyzed the data yet. So, there is no limitation to discuss as we do not yet have an ICC value. We will report the outcome when we know and discuss any corresponding limitation accordingly.

-       I am curious about the choice of 'cannot decide' in the following case: (3) "when the choice fits neither the high- or low-specificity question". Wouldn't this response be rather informative for the research question and worth coding in itself – maybe as exploratory?

Absolutely, we will explore all the data points that receive the designation of "cannot decide" to see if there are any patterns. And we will clearly mark those analyses and findings as exploratory.

8.   Could the authors provide some more context/justification for the following decision: 'Our desired level of precision is that the width of parameter coefficients' 95% HDIs should be equal to or smaller than 16'?

The level of precision we chose is based on Neequaye and Lorson's (2023) findings. We have added the citation; thanks for drawing our attention to this oversight. It is also worth mentioning that the sampling plan was based on joint consideration of resources and previous research (kindly see the discussion specifying ROPE).

Review by anonymous reviewer 2, 14 Aug 2024 22:10

Note: I tried to attach as a separate document, but it wasn't clear if it uploaded or not, so I just pasted here. You can ignore the attached if there.

Thanks for thoroughly engaging with our proposal. We appreciate your attention to detail and the comments assisted us in reexamining our propositions and the underlying reasons.

1A. The scientific validity of the research question(s).

The research questions are scientifically valid, in that they are answerable via research. I had no concerns regarding the ethics. However, I found the theoretical basis to be weak. There was little-to-no reference to the research literature beyond the authors' own previous study (see below for more on this).

We tackle the concern raised here under 1B below to avoid unnecessary repetition.

1B. The logic, rationale, and plausibility of the proposed hypotheses, as applicable.

I found the hypotheses to be appropriately precise, and sufficiently conceivable as to be worthy of investigation. That said, I found the wording of "preference for pragmatic correspondence" to be confusing, as the outcome isn't the participants' preferences. A slight rewording would be helpful.

However, as noted in relation to criteria 1A, I did not find them to be based in any clear theory, or follow from past findings (again, with the obvious exception that the authors were seeking to replicate their prior work and explore the boundaries of those findings). What was missing for me was more explanation of the mechanism. Are the predictions based in theories of memory activation? Or based more on how people interact socially / conversational norms? When considering how to provide answers to questions (a different but related topic) we know people tend to consider both accuracy and informativeness. Is that part of the thinking here as well?

Theoretical rationale: The present research is based on established research and theory in Pragmatics, specifically Relevance theory (i.e., Sperber & Wilson, 1987, 1995) and Gricean Norms (Grice, 1975, 1989). We did not recap this aspect in the present manuscript because we have discussed such rationale elsewhere (i.e., Neequaye & Lorson, 2023) and wanted to avoid repetition—which is why we referred readers to that work. Nonetheless we see the reviewer's point and have included a concise summary to provide a rationale (see pages 3 [last paragraph] to page 4).

The distinction between the present research and related work: As noted above, the present research is based on pragmatics, which is related but different from memory. We have included a paragraph on pages 5 through 6 explaining the distinction. "Theories on memory deal with what makes people (e.g., interviewees) report accurate or inaccurate information. The present research takes a step back to examine how interviewees determine what a question is questioning or asking for. This aspect of a conversation is a pragmatic matter, *yet to be* an issue of memory. One must first engage in some pragmatics—or decipher the objectives of a question—before turning to their memories to craft an answer (see Neequaye & Lorson, 2023 for an extended discussion). Consider the following example. An interviewee can accurately flag that the interviewer wants to know about a bomb design (pragmatic issue), but the interviewee may fail to remember the design because they did not encode that information when they previously encountered the bomb (memory issue)."

Clarifying Core Hypothesis 1: We have now expanded on the hypothesis to clarify the meaning of pragmatic correspondence (see page 9).

New References:

Grice, H. P. (1975). Logic and Conversation. *Speech Acts*, 41–58. https://doi.org/10.1163/9789004368811_003

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.

Sperber, D., & Wilson, D. (1987). Précis of Relevance: Communication and Cognition. *Behavioral and Brain Sciences*, *10*(4), 697–710. https://doi.org/10.1017/S0140525X00055345

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed). Blackwell Publishers.

On page 6 I was a bit thrown by the paragraph starting "Before an interviewer proposes any question, it is reasonable for the interviewees to assume that the elicitation of complete details is the de facto purpose of the interview." This seems to argue against their first core hypothesis. If this is being presented as an alternate theory that didn't fully come across.

Apologies for the confusion. The sentence the reviewer highlights here is, in fact, the Revision Hypothesis: "Question-specificity has no effect when interviewees determine what an interviewer wants to know." The entire research, as noted in the abstract, is to ascertain the robustness of the Core versus Revision hypothesis: "High-specificity questions lead interviewees to focus on particularly relevant details to the exclusion of other information, while low-specificity questions make interviewees focus on a broader range of information items (Mechanism-1)—versus—Interviewees generally assume that interviewers want to know all the information at their disposal, irrespective of question specificity (Mechanism-2)".

We hope the issue is clear and welcome further pointers about potential clarifications.

1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable).

Positives: The rules for data exclusion are stated; it's noted that coders will be kept naïve to hypotheses; the order of scenarios will be randomized; participation in R1 and R2 will happen concurrently; there are attention and manipulation checks; R2 clearly follows up on a potential explanation for findings in the past work while R1 seeks to replicate past work with a similar design; the sample size, power, and sampling plan are acceptable

Thanks for the kind words.

Concerns:
1.   It's not clearly stated that random assignment will be used for the between subjects manipulations (I rather assumed it would be, but it should be stated in the text/procedure).

The between- versus within-subjects designs depend on specific studies. On page 12 (paragraph 3), we note that "Replication 1 will employ a within-subjects design for the question-specificity trials.". An

on page 13 (paragraph 2), we note that "Replication 2. This study will employ an identical protocol to Replication 1, but we will use a between-subjects design for the question-specificity trials. **Participants will undergo five randomized scenarios** in which the interviewer consistently asks either high- or low-specificity questions."

2.   How the manipulation check regarding disposition would be scored wasn't stated, and I did not find it to be obvious. Further, it wasn't clear if failure to answer this "correctly" would result in exclusion or not. Presumably not, as other exclusion criteria were noted.

The scoring of the disposition check is stated in Appendix A (range -1 to +2). This variable will not be used in any exclusions, which is why we did not include it in the exclusion markers. The variable will be used to explore whether interviewees generally understand the implications of the disposition manipulation.

3.   It wasn't clear what the "decision making" manipulation check was (in Exclusion Criteria section) but I assume it refers to that "instructional" manipulation check (in Appendix B).

Yes, the check refers to the instructional manipulation check (IMC) in Appendix B, which is why we added the IMC abbreviation when mentioning it in the exclusion criteria. Apologies for the confusion.

4.   The textbox prompt "The police-contact wants to know if…" seemed odd to me. Specifically, the word "if." That seems like it would generate answer like "…if I know what brand of drugs the gang is selling" as opposed to answer like "…the gang is selling off brand oxy" (I assume the latter is the type of answer the authors' are seeking. I realize the authors piloted this, so I'm willing to defer to them on this point. It just seems very odd to me.

This is a fair concern, and we explored many ways to execute this aspect of research design. The prompt "The police-contact wants to know if…" was the best at ensuring that participants focus on **what the interviewer wants to know**, not what they want to disclose. To reiterate, we are not examining what interviewees eventually disclose so it is crucial ensure that participants remain in the frame of mind where they are hypothesizing about the interviewer's objective. The word "if" is critical in achieving the goal of focusing participants on pragmatics. So a choice like "…if I know what brand of drugs the gang is selling" is exactly what we need in this research because it clearly demonstrates the participant is thinking about an objective: in this case a brand, which will be "off brand oxy".

5.   The "optional" wager question comes out of nowhere – why is it included? It's also unclear what is meant by optional. They have the option to skip it and not indicate what/if they would wager? Or it's optional to wager something (i.e., they have the 0% option)?

We noted on page 11 that participants will provide two confidence ratings. Hence, the wager is an additional confidence rating for robustness. We have added the following text to clarify. "Given that the wager is optional, participants may decide to skip it. This setup provides two extra measures of confidence besides the mandatory confidence rating: (i) whether a participant is confident enough to place a wager and (ii) the extent of confidence as evident in a wager's magnitude."

6.   The authors state that in their past study, and in R1 and R2, pragmatic correspondence was designed to be equivalent to complete details – so high specificity questions specifically request complete details. I thought I understood this, and it made sense to me. But once I got to the Appendices and saw the scenario information and the high and low specificity questions, I had some concerns. In fact, this is my biggest concern regarding the proposed studies.

For the sake of conciseness and to avoid confusion, we will respond to the specific concerns in the order they are raised below.

Let's take the first scenario:

One day after work, on your bus ride home, you recognized one of the KET22 members. You were sitting just behind him, and he was talking on the phone. He tried to be quiet, but you heard him say: "It is better to sell the off-brand green-star oxycodone."
a. Have you discovered the particular brand of narcotics KET22 sells? [HIGH]
b. Have you discovered anything about the gang's narcotics sales lately? [LOW]

I would argue that the first question could be answered with "they are selling off brand drugs" – I don' t how it's seeking complete information. In contrast, the second, low specificity, question seems to be seeking complete information.

Thanks for raising this issue. We respectfully disagree with the conclusions drawn here. Asking for the **particular brand of narcotics** the gang sells objectively requests all the relevant information in the scenario. The answer proposed by the reviewer as the most pragmatic answer is incomplete given the question. "off brand drugs" is not the complete information "off-brand **green-star** oxycodone" is the particular brand of narcotics the gang is selling. One cannot say "green-star" without also saying "green-star" refers to an oxycodone, and the information is still incomplete unless one adds that "green-star" is an "off brand". Yes, a person could theoretically think the reviewer's proposed answer "off brand drugs" is the question's goal—but, objectively speaking, that choice is not the goal of the question.

The low-specificity question asking for "**anything** about the gang's narcotics sales lately" does not request complete information. The word **anything does not request specific items;** it asks for whatever information a person might possess. And that possibility introduces a range of reasonable responses as listed below. Item-i through -iii are objectively accurate answers to the question asking for **anything** item-i and -ii are not accurate answers to the question "Have you discovered the particular brand of narcotics KET22 sells?"

   i.   They sell oxycodone.
  ii.   They sell green-star oxycodone.
 iii.   They sell the off-brand green-star oxycodone.

Below is the English dictionary definition for the word "anything". Anything refers to a thing, **no matter what**. Anything is used **to indicate range**. Emphasis does not apply here because we are dealing with questions.

anything | ˈɛnɪθɪŋ |

pronoun [usually with negative or in questions]

used to refer to a thing, no matter what: *nobody was saying anything* | *have you found anything?* | *he inquired whether there was anything he could do.*

- [without negative] **used for emphasis:** *I was ready for anything.*

- **used to indicate a range:** *he trains anything from seven to eight hours a day.*

This can be contrasted with the second scenario, where the high specificity question is clearly asking for full details (and questions like this are in line with what I was expecting):

You always come to work earlier than your colleagues because you supervise the cleaners. You've realized that the KET22 gangsters usually arrive shortly after you in a blue Nissan Qashqai. By paying more attention, you've memorized the license plate number: FBT038.
a. Do you know the full details about the vehicle the KET22 gangsters usually arrive in at the park? [HIGH]
b. Do you have any information about KET22's transportation in the park? [LOW]
Given that the reviewer, does not have an issue with this scenario we will move on to the next example.

Turning to third scenario, this seems more like the first, where a very specific piece of information is requested by the high specificity question: "EXIT 7F." In contrast, the low specificity question seems to invite any and all information.

Lately, you have noticed a particular spot at the park where the KET22 gangsters deal drugs in the evenings. The spot is one of the park's exits, EXIT 7F. All the exits are located at different edges of the park, but 7F is rather discreet.
a. Have you spotted the exact location at the park where KET22 deals drugs? [HIGH]
b. Have you spotted anything about where KET22 deals drugs? [LOW]

Yes, this scenario is like the first one and merits a response like the one we provided. In this scenario, when one says "EXIT 7F." They must necessarily include all the other information that it is at the "edge of the park" and "discreet". The information bit "EXIT 7F." is not meaningful by itself unless one adds the contextual description

As the reviewer points out the low-specificity question invites "any and all information"—which is exactly the manipulation we are going for. The hypothesis is that low-specificity questions would invite any information about a scenario, whether it is complete or not.

I will not go through each scenario and its corresponding question. These differences need to be addressed. Either the authors think that asking for a very specific detail will elicit all relevant information (i.e., scenarios 1 and 3), or they think that asking for full information will elicit all

relevant information (i.e., scenario 2). But those seem like very different things, and either way, this should be clear in the manuscript, and the scenarios should be consistent in how this is approached.

Yes, we specifically designed the scenarios and the corresponding high- versus low-specificity questions such that the specific details requested by high-specificity questions will necessarily come with all the other details; otherwise, the specific detail will not be intelligible. We hope that the clarifications discussed about Scenarios 1 and 3 are now clear. They are similar to Scenario 2.

If one examines all the scenarios and questions, one will notice this pattern is consistent. Regardless of the framing, all the low-specificity questions are defined by the pronoun "**any**" to introduce broadness such that one could think the interviewer wanted to know a range of things, not necessarily complete information. And high-specificity questions necessarily require all the information in a scenario—we disagree with the implications the reviewer draws about the pragmatics of Scenarios 1 and 3.

Indeed, our previous research (Neequaye & Lorson, 2023, same scenarios as current study) found that participants were more confident that they had identified the interviewer's objective when asked high- versus low-specificity questions. So, at the very least, we can be confident that the **low-specificity questions are generally more broad in their perceived objectives** than the high-specificity questions.

 I also have concerns with the fact that scenarios contain such limited information. To me this makes it reasonably easy for all participants to choose to provide all information. Unless I'm missing something, it seems like participants don't even need to be presented with any information. They could just be presented with questions and ask what information they think the interviewer would want them to find out. This is less leading as there are many potential options, not a couple details.

Thanks for raising this concern, but we respectfully disagree with the implications and conclusions drawn here—at least given the current findings and progress in this research program.

The research is **not** about how interviewees determine what interviewers **would** want to know—with the objective of **finding out that information** (as the reviewer writes). Such a process might recruit mental processes, unlike our theory—we simply do not know and have not theorized about such a case.

The present research is about how interviewees determine what interviewers want to know, **given question-specificity and the information the interviewee already possesses**. The research is about how interviewees slice and dice the information they already possess—given an interviewer's question (whether the question is leading or not), which is different from receiving a question and then going on to find answers. Going on to find information is a matter of fishing for something new, not splicing what one already knows.

We believe our approach is more robust and provides a severe test of our Core Hypothesis. Suppose—even with limited information—interviewees still focus on some details to the exclusion of others as a function of question-specificity. That finding would strongly support our hypothesis. Note that we are currently testing a within-subjects versus a between-subjects paradigm.

As we have planned and stated in the manuscript, suppose the findings (i.e., Replications 1 and 2) indicate that interviewees focus on all the information regardless of question-specificity. Then, we

would have a warrant to explore expanded scenarios (Replication 3) to examine whether such a finding is due to limited information in a scenario.

"We will specify the exact design and conduct Replication 3 after analyzing Replications 1 and 2. This sequential approach will assist in determining the appropriate stimulus material that can best address the suggested limitations of NL (Study 1). (p. 9 current manuscript)"

Indeed, some of the low specificity questions are so vague that there are a huge number of details that an interviewee might suggest the interviewer was interested in, if they were not confined to 2 or 3 pieces of information. (e.g., "Have you discovered anything about the gang's narcotics sales lately?: This could be getting at whether sales good or bad; is one product selling better than another; sales are initiated via text messages; sales are primarily conducted by person X and person Y). As the proposed studies are more of an initial preliminary test, this is less of a concern than the previous point I made, but something to consider moving forward.

Thanks for raising this point. Indeed, such issues are in the pipeline for this research program when we have more data on how people deal with the pragmatics of low-specificity questions. Then we may generate new ideas about the effect of low-specificity questions

Even so, as the reviewer just said, the scenarios in this research contain limited information. So, per Gricean norms of relevance, the broadness of low-specificity questions is still limited to the contents of a specific scenario. The goodness or badness of sales in the example the reviewer offers is not included to the scenario (i.e., the information the interviewee already possesses). Our pilot studies (also available on OSF) show that participants don't generate information outside of a given scenario's contents.

 Note: I must apologize, but I do not feel I have the relevant statistical expertise to provide a helpful evaluation of the statistical analyses proposed. (Nothing struck me as incorrect or concerning.)

1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses.
Thanks for this comment. We tried to describe the studies as best as possible without an excessive word count, which is why we have made the materials completely open. The Qualtrics QSF file can be used to reproduce the procedure verbatim, preventing flexibility in the procedure.

I appreciate the authors inclusion of the study materials; these were very helpful. That said, without providing the materials in the appendices I don't think the methods would be sufficiently clear.

A few places I was still a bit unclear, even with the appendices:
-The presentation of the control questions was not entirely clear to me. This should be better incorporated into the procedure.

Thanks for noting this issue. The control questions were randomly distributed among the scenarios in each study. We have now included this aspect of the design in the manuscript (p. 14).

-The "disposition" manipulation was introduced in such a way on page 8 that I didn't realize it was a

manipulation. I thought at first that participants could choose their disposition (also the term disposition was not introduced at that time, so it took me by surprise later).

Apologies for this oversight. We have now edited this aspect of the procedure overview to clarify. "The informant role will be manipulated so that participants will take on one of the following dispositions: cooperative, semi-cooperative, or resistant when engaging with the investigator. (p. 9)"

-Coding: I was quite confused here. I think part of the is that perhaps where the authors said "choice' they really intended to say "response." If that is the case, I think I understand what the authors are proposing, but, it doesn't quite make sense to me. I think this is related to my major concern above. My understanding is that the authors are proposing that coder should code more "complete" answer on the "high specificity" side of the scale. But that assumes that the high specificity questions really were in fact seeking complete details. Which, as I note above, I don't think they (all) are.

This is a fair point. We agree that the word 'response' is more intuitive. Previously, the word response has led other reviewers to assume the research was examining disclosure instead of pragmatics. For that reason we used the word 'choice'. We have now replaced choice with response.

1E. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s).

As noted above, there are manipulation checks included. Overall I don't have concerns with this criteria beyond anything I have already mentioned.

Many thanks again for the comments.