**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).

**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?

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.

**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

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).
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.
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).
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.
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)?
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.

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.

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]*

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]*

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.

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. 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.

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.**

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.

-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).

-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.

**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.