**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.
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?
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.
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?
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?
6. Page 25: The authors say there were no anomalies in the data. How do they operationalize that?
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.
   1. 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?
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.
9. I assume the planned study will be registered, with its data, code, and materials made openly available?

I always sign my reviews,

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