22nd March 2022

PCI Registered Reports #154: Exploring How Members of Illicit Networks Navigate Investigative Interviews

Dear Zoltan,

Many thanks for the opportunity to revise and resubmit the captioned registered report. We also thank both reviewers for investing the time to assist us in improving the registered report. We have now revised the registered report according to the concerns. The revised version was proofed to eliminate ambiguity and confusing sentences. The clarified manuscript addresses some of the issues the reviewers raise by shedding light on possible misunderstandings. Additionally, this letter contains a detailed point-by-point response to their concerns.

We bring some design modifications/issues to your attention and seek your approval/decision before we possibly commence data collection. For conciseness, the issues are noted briefly below but described in detail in our responses to the comments.

- Based on Prof. Hope's comments, we will now include an exploratory measure asking about possible roles. The additional question asks participants if they decided on any specific roles in their group in preparation for the interviews.

- Based on Prof. Ormerod's concern about collusion and information sharing, we will include features to ensure robustness. We will include reminders in Phase 1 and 2 telling participants that communication between group members is not allowed during the interview (see appendix). Before the interview, each participant will be required to confirm that they will not communicate with other group members. Additionally, we will include a check after the interview, asking participants to confirm whether they communicated with any group member during their interview (see appendix). Suppose any participant flouts the instructions: we will conduct a robustness check by presenting the results with and without those participants. We have added this aspect to the manuscript (page 13)

- Prof. Ormerod was concerned about the possibility of measuring affiliation before the interview phase biasing the data. However, it is not immediately apparent to us how measuring affiliation before the interview could bias the data as the measure does not aim to manipulate affiliation. Pending your decision, we have planned to leave the affiliation measure unchanged.

- Prof. Ormerod also commented on the maximum time we will allot to participants for planning. We chose 20minutes because of the length of the background story and the compensation we can offer. We are prepared to increase the maximum planning time to 30 minutes if the editorial team insists.

- For documentation and replication's sake, we will upload a Qualtrics (qsf) file corresponding to the Phase 2 procedure before collecting any data. We believe this addition will eliminate any flexibility and ambiguity in future replications.

Next follows our point-by-point response to the reviewer comments. To avoid confusion or misquoting, we have included the reviewer comments verbatim (in black font) and our response follows in red font.

Sincerely
Authors

Table of Contents

Comments and responses: Ormerod T.C.

Summary: A single experimental simulation of intelligence gathering from interviews with members of a mock network is proposed, as a test of the disclosure-outcomes management (DOM) model, and to inform practitioners on appropriate interview strategies for intelligence gathering from network members. In phase 1, participants interact in groups of 6 to plan their intended information revelation. In phase 2, individuals go through an online 'interview' comprising a sequence of interviews and selections from a drop-down menu. The extent to which their information sharing is accounted for by mixed effects random models of individual and group presents their approach to testing whether individuals make their own disclosure decisions or are guided by the group plan…. I think (-:

Recommendation: This is a fascinating area to explore, and an ingenious approach. The study has been designed with precision and the analysis plan looks sound. It is a very complex design indeed, and I am not 100% sure I fully understand it. My main concern is about the practical relevance of the research design to the real task of intelligence gathering. I acknowledge it is exceptionally difficult to study (see, for example, our own struggles: Dando & Ormerod, 2019 – Journal of Experimental Psychology: General). But I do worry that the design is so convoluted and potentially confounded, and that the task set is so far from anything that would realistically be encountered in an intelligence-gathering setting, that the efforts in data collection may end up providing more noise than signal. I raise a set of issues below that I think need addressing. I do hope a version of this study can be accepted, but I also think the issues below require attention or a robust dismissal before this pre-registration can be accepted.

Many thanks for taking the time to review our proposal. We are especially thankful for the level of engagement. We do not take for granted that you invested quality time in this review; your efforts are much appreciated. In what follows we respond to all the comments and clarify the issues of concern.

Hypotheses:

My understanding of the predictions of DOM are summarised in this table:

	Low benefit	High benefit
Low risk	"low stakes" – no reveal	"Unguarded" - reveal
High risk	"Guarded" – no reveal	"High stakes" – maybe reveal

Re. Q1 – The DOM predictions seem fairly facile to me. What theory would predict otherwise? I think this needs to be addressed explicitly for the value of the study to be justifiable.

To be concise, we described only the relevant aspects of the DOM model in the current manuscript. Another paper describes the theory and its contribution in more detail (see Neequaye et al.,

2021). We are happy to include more detail in the current manuscript if the editorial team insists. In what follows, we reiterate the contributions of the DOM model. That perceived costs and benefits would determine what people disclose is consistent with what anyone would expect. The DOM model contributes nuance to how the research field conceptualizes disclosure in intelligence interviews.

First, the model brings the importance of the self-interest dilemma to light. To mimic real intelligence settings, intelligence interviewing experiments usually incorporate a goal conflict or what we call a self-interest dilemma in their research designs (Dawson, 2015; Neequaye, 2018; Oleszkiewicz, 2016). Participants are instructed not to share too little information because assisting the interviewer is necessary for some personal benefit. The participants are also told not to share too much information since they are to imagine having strong ties to a mock terror or criminal group, for example. The two-fold instruction has proven useful; all the participants usually disclose some but not all the information they possess (see review by Luke, 2021). But it remains unclear why interviewees elect to reveal the information they do. The DOM model explains that the specific elements of the self-interest dilemma determine the characteristics of the different pieces of information the interviewee holds. That conceptualization engenders the information-types, which offer different predictions on what an interviewee is likely to disclose. Existing research focuses heavily on the amount of information interviewees disclose, which says little about the underlying mechanisms of disclosure. As Prof. Hope remarks, all illicit information is not equal. The DOM model brings Prof. Hope's remarks to light.

Second, the DOM model moves the field toward considering pieces of information as the primary unit of analysis of disclosure, not just the amount of disclosure. This shift offers many conceptual and analytic insights. By focusing on lower-level units, we can begin to offer idiographic predictions—that is, hypotheses pertaining to the behavior of specific individuals rather than averages about the amount of disclosure. An interviewer who plans an interview taking an interviewee's self-interest dilemma and consequent information-types into account can be more judicious. They can

formulate strategies to elicit specific pieces of information, not just more information. Additionally, the DOM model allows the field to consider how person-level variables (e.g., personality) and other factors, in the present case group planning, might interact with the properties of information to influence disclosure or non-disclosure. Such considerations may lead to more sophisticated predictions about behavior.

P3 "When interviewing someone about their network—and the focus is solely on the network—to what extent does network membership predict the *type of information* the interviewee will *choose* to reveal. Are two or more different people from the same network likely to disclose similar kinds of information? Or does disclosure in this context better resemble individuals independently managing the potential outcomes of revealing information?" Aren't all the 'networks' essentially identical, so how could this prediction be tested? Surely it would be a much more powerful study if the underlying structure of networks was manipulated (e.g., a hierarchy vs. a heterarchy) so that network influences could be detected with precision?

Models 3, 4, 5, as laid out in the analysis plan, will examine the question flagged here. Illicit networks might be alike in that they all engage in nefarious activities. However, different networks, and the members therein, might approach their illicit affairs differently. For example, two narcotic dealing criminal gangs can use different modus operandi to traffic narcotics and evade law enforcement despite their alikeness as criminal gangs. Hence, it is useful to know whether people from the same network behave similarly. Suppose our findings reveal that members of the same network manage the costs and benefits of disclosure in the same way. Then we might advise interviewers that a strategy to elicit information from one member could be effective when interviewing another member from the same network. Suppose the results also indicate that individual decision-making exerts a strong influence. Then we might advise the interviewer to be

mindful of how other members of the network behave but not take the tendencies of the particular member undergoing the interview for granted.

We agree that a study that manipulates something specific about a network will be insightful. However, our goal at this <u>initial</u> stage of research is to examine a <u>basic or generic</u> form, which might characterize some networks: a group (not a dyad) consisting of a flat hierarchy with a common illicit goal contending a dilemma on what to disclose in an investigative interview to further that goal. Note that there is hardly any research in this area, making it prudent to begin at a basic level. Our findings may then be used to determine procedures, effects sizes, etc., for studies manipulating different aspects of a network. Indeed, different things could characterize a network: for example, (1) whether the dilemma focuses on the individual members or the network as a whole, (2) Hierarchy (3) Loyalty (4) Familial ties (5) centralized v decentralized (6) homogeneous v heterogeneous structure. It will be too challenging for one study to perfectly capture all the forms networks might take. We elected to begin this line of work in a manageable way and build on the present basic design once we gather more insights. On page 4, we include an explicit note to clarify this issue.

P5" Such disclosure is not immediately beneficial to navigating the self-interest dilemma but carries potential costs. "

I thought the point of low-stakes information was that it doesn't carry potential costs. The DOM model does not therefore predict low levels of disclosure – one would need a theory of verbal manoeuvring such as Interpersonal control theory to derive this prediction.

There has been a slight misunderstanding here. Our conception of low stakes information is that <u>the perceived costs and benefits of disclosure are low</u>, not that they carry no potential costs. Thus, we predict that people will typically refrain from disclosing low-stakes information because

disclosing such information is not worth taking on possible costs when the potential benefits are low. Indeed, our earlier study on the DOM model aligns with this prediction (see Neequaye et al., 2021). That said, we agree that more theorizing might be needed to determine influences that might stifle of enhance the disclosure of low-stakes information. But deciphering elicitation techniques is not our goal in the present work. We are after the mechanisms underlying disclosure. The results would then help theorize about intervening factors that influence perceived costs and benefits.

P6 – "result a" vs "result b" – The comparison between results a and b seems to boil down to this: if this research tells us that everyone in a network will be influenced by the same cost/benefit judgements, then they will all say the same thing in interview, so you can interview them all in the same way. If not, they won't, so you have to treat each individual differently. I don't understand how this informs an interview questioning approach. The detective identifies from another source (a tip off) what would constitute a high stakes piece of information. She/he then interviews one member network who reveals that high stakes piece of information. So the detective thinks 'right, I'll repeat the same interview approach because that revealed the high stakes information'.   But a) they already knew the high stakes information, and b) they had it confirmed in the first interview, c) knowing what is high or low stakes isn't the thing that is helping the detective: what is helping is the interview method they chose for the first interview. Perhaps most importantly, d) the idea that one can draw any kind of generality from this research to interviewing members of a real network, where there are different roles, hierarchical positions, motivations, fears, intellectual competencies, personalities, etc. among members, seems wrong. To advise any detective to ignore individual characteristics in planning an interview approach would seem to be very poor advice indeed.

First, we must clarify that "result-a" and "result-b" are not competing hypotheses. As noted on page 6, those projected results are subsumed under Question 2: To what extent do individual decision-making and network membership independently and jointly predict the kinds of information people disclose about their network? Thus, results-a and -b will complement each other to assist us in understanding the mechanisms of disclosure when network members undergo interviews. In the response before here, we discuss the implications

Additionally, a finding that members of the same network are likely influenced by similar cost/benefit judgments does NOT mean they will necessarily say the same thing. That finding suggests they manage costs and benefits in the same way—for example, that members of this network tend to take risks by disclosing high-stakes information when the interviewer presents any benefit of disclosure. Consequently, an interviewing approach that elicits a type of information from one member might elicit the same type of information from another member.

This theorizing highlights the contribution of nuance the DOM model brings. An interviewee's "cooperativeness", via disclosing information, varies within the individual; and the variation occurs between different pieces of information the interviewee holds. Therefore, the DOM model's unit of analysis is pieces of information nested within interviewees. Any given network member might hold multiple items of information, which we capture in our research design with multiple items of the FOUR information-types. The fact that an interviewer has managed to elicit one piece of high-stakes information from, say, Network Member-A does not mean the interviewer has exhausted all possible high-stakes information on a topic. Yes, the interviewer might confirm that high-stakes information when interviewing Network Member-B. But there could be DIFFERENT high-stakes information to elicit, not just one. And now that the

interviewer has "established what works", she could use what she now knows is an effective interviewing approach to elicit DIFFERENT high-stakes information from Member-B, -C, and so on.

Knowing what information (and topics) might present as low-stakes, high-stakes, guarded, or unguarded information helps the interviewer better predict how interviewees might react when an attempt is made to elicit the information. Then the interviewer can determine what interviewing approach might be effective on a first attempt and adapt over multiple interviews. An interviewer can make those predictions by considering individual and network characteristics. As noted earlier, we are examining how individual decision-making and network membership independently and jointly predict the kinds of information people disclose about their network. We definitely do <u>not</u> intend to advise interviewers to ignore anything but rather to advise on the extent of influence individual and network characteristics might exert.

We do acknowledge the limitation about the generality of the present research to all the different forms a network might take in the real-world. As noted in the response before here, this research is a start focusing on networks in the broadest form—a group of people (not a dyad) with a common illicit goal contending a dilemma on what to disclose in an investigative interview to further their goal.

P7." We aim to recruit already acquainted participants—for example, friends or co-workers—to serve as a network: typically, network members are not complete strangers. This design choice allows us to commence test sessions without needing to induce familiarity between participants"

This decision imposes a limitation on the generalisability of the results. a) Doesn't this bias the results towards intra-network agreement (i.e., a confirmation bias on the predicted results)? B) It doesn't necessarily reflect the reality of networks (e.g., Al Quida is a brand not a network of friends),

and c) it introduces huge sources of variability (notably experience and trust) that do not form part of

DOM theory.

    We respectfully disagree with the assertion that our design choice of using acquainted people

will bring unnecessary noise to the data. We purposely included the Inclusion of Other in the Self (IOS)

Scale (Aron et al., 1992) to measure variability in affiliation. Variability in affiliation will arise even if we

use strangers because we would have to induce some form of affiliation to capture a basic essence of

networks. Members of networks are usually acquainted and are affiliated somehow (Krebs, 2002;

Ouellet et al., 2017). And we hope to capture this fundamental aspect of networks in the research

design.

    Moreover, our research questions do not rely on intra-network agreements to confirm any

predictions. We are examining the extent to which group planning and individual decision-making jointly

and independently influence disclosure decisions.

    Finally, the DOM theory does not theorize about networks but rather information disclosure. As

noted in Question 1, under The Present Research Section, the proposed research will examine how well

DOM generalizes to the context of illicit networks. Here we examine the extent to which dilemmas at

the level of an illicit network generate the information-types the model predicts. An intra-network

agreement could just as well lead participants to ignore the cost-benefit manipulations.

P7. They propose 20 networks of 6 individuals in each. Is this 20 x 6 = 120 or just 20 data points?

Note on P3 they state "We will explore how perceived disclosure outcomes, namely, the projected

costs and benefits, affect what networks choose to reveal. The study will focus on disclosure

pertaining to the network as a whole, not about the individual being interviewed, per se" So is it

networks or individuals? I think its individuals not networks, so this quote from p.3 really needs to

be adjusted.

A misunderstanding has occurred here. We will examine the influence of network membership (viz., the 20 networks) AND individual decision-making (viz., the 120 participants). Participants will make 48 decisions resulting in 5760 data points. So, it is NOT networks or individuals; we will examine the influence of being a part of a particular network AND the influence of the individuals in that network.

Secondly, as noted on page 3, we are examining a case where an investigation focuses on aspects of the network, not the specific individuals in the network. Therefore, the study examines the extent to which group and individual decision-making predict the management of information disclosure <u>about a network</u>, not information disclosure about the specific network members undergoing an interview. On page 3, we describe another type of investigation, which is not our goal in the present study.

To avoid potential future misunderstandings, we have included a clarifying sentence on page 3. "Our quest is to explore the extent to which group and individual decision-making predict the management of information disclosure about a network, not information disclosure about the network members undergoing an interview."

P9 - "They will execute their decisions by selecting what to disclose from a list of possible information items"

Doesn't this create a massive demand characteristic to disclose (as opposed, say, to a 'no comment' interview, a much more realistic outcome of an intelligence interview)?

We do not believe that selecting what to disclose from a list of items creates demand characteristics. Our previous use of this protocol indicates that participants select what they want to disclose. Moreover, we explicitly instruct participants that they can decide to say nothing by selecting none of the items. If the editorial team insists, <u>we can include a "no comment" item</u> in the choice-list.

However, as noted we have had no issue with this procedure in the previous execution (see Neequaye et al., 2021).

P9. This took a lot of understanding, and I am not sure I do understand it. At first I thought "Doesn't the 'probability' information effectively tell participants what they can and cannot reveal?

We appreciate the concern raised here, and in what follows, we will try to explain how the design does, in fact, capture the dynamics of strategic disclosure in intelligence interviews. As noted in the methods, we use an incentive-compatible procedure, which studies have found leads participants to indicate their true preferences (Wertenbroch & Skiera, 2002). The probabilities of the benefit and costs probabilities align with the background story. They mirror the dilemma in the background story and investigation regarding what <u>might</u> be dangerous or safe to disclose. But one cannot predict such costs and benefits with certainty. So, the probabilities do NOT tell participants what to disclose or withhold. The probabilities (which align with the dilemma) mimic the situation whereby one anticipates probable consequences of disclosing an information item (see, e.g., Soufan, 2011; Toliver, 1997).

So, they will simply maximise release of low risk and minimise high risk".

So, <u>no,</u> it is not simply minimizing and maximizing low and high risk; the study will examine whether the 4-prong combination of costs and benefits (not just costs) indeed map on to our predictions of information-types.

But having got the whole way through, it seems that the probability information is only marginally predictive of the actual gains and losses. That said, it seems like the participant only discovers that once they have had the second video (i.e., the first trial). So, in all likelihood, participants are going to treat these probabilities as if they are veridical rather than pseudo-random in their informativeness.

We are happy to discuss what happened during the test run, but the probabilities are INDEED predictive of actual gains and losses (see appendix for the code we will use). "Disclosing a given piece of information will bring a random outcome based on the provided probabilities. If the sum of the

probabilities of the positive and negative outcomes does not sum to 100%, the remainder will represent neither a positive or negative outcome (p. 11)". So, a given disclosure could bring a positive, negative, or null outcome. These consequences align with what might happen in the real world. As noted before, we have implemented this procedure successfully in a previous study.

Providing the % safety information during the interview seems a bit odd – presumably the group will discuss what they think is safe before, and then this % info comes along that may override or confirm the group discussion… how does that work? Is the idea that the % info should conflict with the group decision? To be realistic, the networks ought to decide what is high and low stakes, and the individuals when interviewed should be left to make their own judgements on the basis of what was planned by the network, and not be told what is high and low stakes by the interviewer.   The proposed approach seems to take the study farther away from anything that has practical validity.

We respectfully disagree with the characterization presented here. First, it should be noted that the interviewer does not tell the participants what the information-types are. The interviewer poses a direct question about a topic, and participants choose what to disclose in response. As we noted before, the percentages mirror the possible costs and benefits, namely, the dilemma in the background story. The results will reflect any severe conflicts between how the networks perceive the story and the percentages.

Allowing participants to determine the information-types will be a test of the contents of the background story, not a test of the DOM theory. By providing the probabilities, we maximize internal validity and test the DOM model's tenets. If our predictions of the information-types do not pan out in this study, then the model has low verisimilitude in the context of networks. The present study is an initial examination, and we prioritize experimental control and internal validity; we do appreciate the concern about ecological validity. However, as noted in the manuscript (under internal and ecological validity), our design addresses the essential research question. To what extent do perceived outcomes

influence what network members choose to disclose? The story and the corresponding cost-benefits probabilities, as well as the consequences, manipulate the outcomes of disclosure to examine what participants choose to disclose.

P9 – "We will tell them that they can double their endowment in the best case, and in the worst case, they can lose the entire endowment."

There seem to be competing dynamics here: on the one hand, the study is looking at group cohesiveness and information disclosure, bit on the other hand it seems to be a behavioural economics study looking at how well individuals can maximise profit and minimise losses. I worry that this financial incentive creates an entirely different set of task goals to the ones the research team think they are studying. All the stuff about keeping/losing licences, stopping/continuing investigations and disclosing some information to the interviewer to appear cooperative is made irrelevant, or at very least is confounded, by this financial incentive.

We appreciate the concern here. However, we believe our incentive-compatible procedure improves on the existing research designs by making the consequences of disclosure tangible, not merely imagined. Most existing studies simply ask participants to strike a balance between disclosing too much or too little information. Or they ask participants to imagine the consequences of disclosure. Our procedure makes the disclosure of any given piece of information tangible, allowing us to measure not just the amount of disclosure but manipulate items of information.

Additionally, the financial incentive's tangibility does NOT confound the disclosure decisions; the incentive makes the consequences real as opposed to imagined. In the real world, interviewees disclose and withhold information because of SOMETHING akin to perceived costs and benefits (see, e.g., Soufan, 2011; Toliver, 1997). Here, we make that SOMETHING tangible by weaving the financial incentive into the story and dilemma. The alternative is to ask participants to simply imagine the dilemma's consequences, which we believe is weaker than the incentive-compatible procedure. The

gains and losses of the financial incentive bring to life the benefits an interviewee might want to gain (e.g., protecting a business license) or costs to avoid (an investigator shutting down the network's business).

P10. Is 20 mins planning enough? Where does this figure come from?

We determined the 20 mins planning time based on the length of the background story and the amount of compensation we can pay participants to invest their time in the proposed research. If the editorial team insists, we could expand the planning time to 30 mins.

P11. "In truth, each participant will be told that the investigators have called them in for an interview."

 If the participants are members of an existing friendship group interviewed over Zoom, how will the researchers avoid collusion and information sharing?

Many thanks for bringing this issue to our attention, we appreciate that you spotted this concern. To ensure robustness, we will include reminders in Phase 1 and 2 telling participants that communication between group members is not allowed during the interview (see appendix). Before the interview, each participant will be required to confirm that they will not communicate with other group members. Additionally, we will include a check after the interview, asking participants to confirm whether they communicated with any group member during their interview (see appendix). Suppose any participant flout the instructions: we will conduct a robustness check by presenting the results with and without those participants. We have added this aspect to the manuscript (page 13)

Furthermore, the research procedure and design address other potential issues on collusion in several ways. First, participants will participate in the same zoom meeting using separate devices.

Second, the interview will <u>not</u> occur over zoom but rather a Qualtrics link that takes each participant to a <u>separate</u> interview; no information sharing can occur. Third, the different topics and the information items therein will be presented in random order. Fourth, we will create three different randomization codings for earnings and losses and present the three possible earning and losses also in a random order such that sharing information among the group will be futile.

P.11 Should level of affiliation questionnaire come after the interview not before, so as not to bias the data?

We will measure affiliation before and after the interview. It is not immediately apparent to us how measuring affiliation before the interview could bias the data as the measure does not aim to manipulate affiliation. We are open discussing the issue as this aspect of the study is largely exploratory.

p.11 Surely presenting % safe and %dangerous estimates is two views of the same measure If its 10% safe, doesn't that necessarily mean its 90% dangerous? This scale is not the same thing as indicating costs and benefits: the cost (e.g., in terms of risk of incrimination) can be entirely independent of the benefits (e.g., showing cooperation with the interviewer), whereas safe vs. dangerous measures are clearly not independent.

We will resolve this issue by presenting the estimates as %beneficial and %dangerous. It is worth mentioning however that we have used the labels safe and dangerous in the without incident (see Neequaye et al., 2021).

p.11. I don't really understand the bit about "disclosing a piece of information will bring about a random outcome" – I am guessing that it means when a participant gives something away, they are told how much cost/benefit in terms of cash that generated. But its not very clear. Having been through the Qualtrics link, I see now that the link between the %s selected and the fiscal outcome is what is randomised. This is really odd for the participant. For example, on trial 1, I selected 15%/15% safe/dangerous figures and ended up with a net loss of 2.5 SEK. On trial 2 I went for 50%/50% and

ended with a net gain of 5SEK. However, I don't think I can have understood this correctly, since on p. 14 the researchers state that the gains and losses with each combination are non-random. Whatever the meaning of this, I worry that there will be unintended carryover between trials on the basis of this trial-by-trial feedback that will interfere with any measure of the influence of network-level planning.

By *random outcome*, we mean to reiterate that participants cannot determine with certainty whether their gains will be beneficial or costly despite the percentage probabilities. Note that the complete sentence is: Disclosing a given piece of information will bring a random outcome <u>based on the provided probabilities</u>. So, in this iteration, the reviewer going for low stakes disclosure ended up attracting a cost, and the reviewer was lucky with disclosing high-stakes information. This *uncertainty* is the very issue the design aims to capture in keeping with how things might go in the real world.

The probabilities are "anticipatory guides", just like an interviewee or the network in the real world might anticipate based on their dilemma that an information item might be beneficial or costly to disclose. However, the interviewee cannot be certain about the estimate until they make the disclosure. The feedback in our design generally mimics such positive and negative consequences, and people typically adjust their behaviors based on consequences. As noted before, our design brings consequences to life instead of participants imagining that a disclosure would have been costly. So, yes, participants might adjust their behaviors or decide to stick with what their network planned (assuming the network's plans conflict with the feedback). But this aspect is not something we want to eliminate because that scenario is what might happen in the real world. Networks cannot predict with certainty the consequences of disclosure in upcoming interviews. Network members undergoing an interview might encounter that network's plan is beneficial, costly, or a mixture of the two. The study examines members of Illicit Networks navigate such costs and benefits. Our question being: to what extent do group planning and individual decision-making independently and jointly affect that navigation?

p.13 – what is a 'memory check' in this context? It needs to be specified clearly, since it might confound the underlying incentive/task structure.

We intend for the memory checks to assist us flag and exclude the data of inattentive participants. The possibility to earn additional compensation for the memory checks was to incentivize active participation. But we do appreciate the possibility for confusion with the main task, so we will now eliminate earnings for passing the memory check (see page 13).

p.13 – what is described as an 'interview' is not in any meaningful sense an intelligence interview. Participants are simply asked to select from a list their responses to three video segments. This lack of realism is highly limiting for the generalisability of the results. I appreciate the researchers have considered this and have made a case for this being a starting point, but I do wonder whether it really tells us anything meaningful about information disclosure during intelligence interviews rather than simply information disclosure, per se.

 We appreciate the concern here and as the reviewer notes, we have been explicit that the present work is a starting point. The goal is to exercise maximum control until we have the experience and the necessary infrastructure for designs including a verbal interview where the nuances of interpersonal communication can be studies and/or manipulated. In any case, as the reviewer notes, the study will speak to information disclosure, disclosure is the crux of intelligence and investigative interviews in general.

NB. Shouldn't the relationship with Prisoners dilemma/game theory be made explicit?

To be concise, we describe the pertinent aspects of the DOM theory for the present research. In Neequaye et al. (2021), we describe the theory in full detail. However, we have now made the connection with existing work on strategic interaction (see page 4).

Comments and responses: Lorraine Hope

This Stage 1 Registered Report (RR) outlines a novel methodology to explore an interesting and neglected question in the field of information elicitation. Specifically, the experiment outlined seeks to examine how several individuals who are collectively members of an illicit network behave in terms of information disclosed when placed in an information management dilemma. This is an important applied question with implications for the investigation of organised crime, terrorism, and other serious/complex crimes. However, it is also of significant scientific merit as our field rarely takes into account the more complex social relationships and dependencies that likely drive interviewee output in such contexts. As such, scientific insights to date are limited by this singular focus and, as such, we have no real insights into the impact of this limited ecological validity. As such, I think there is also a strong scientific rationale for the general approach outlined in the RR. I also find the methodology outlined innovative with respect to tackling a challenging question.

I have a number of comments and questions for the authors. First, I would like to note that all of my comments/questions are offered in the spirit of good faith and the goal of providing constructive feedback. Given than this is an RR, I'm also trying to frame my comments as an open research development discussion – as opposed to a review focused on potential flaws (which is why I've included some references to our own recent experiences with similar paradigms – apologies if this is in any way inappropriate). I should also note that I have completed this set of comments under significant time pressure so I expect other thoughts will come to mind later which I'll be happy to forward if they seem sensible as I'm excited by this work and genuinely interested in seeing it come to fruition.
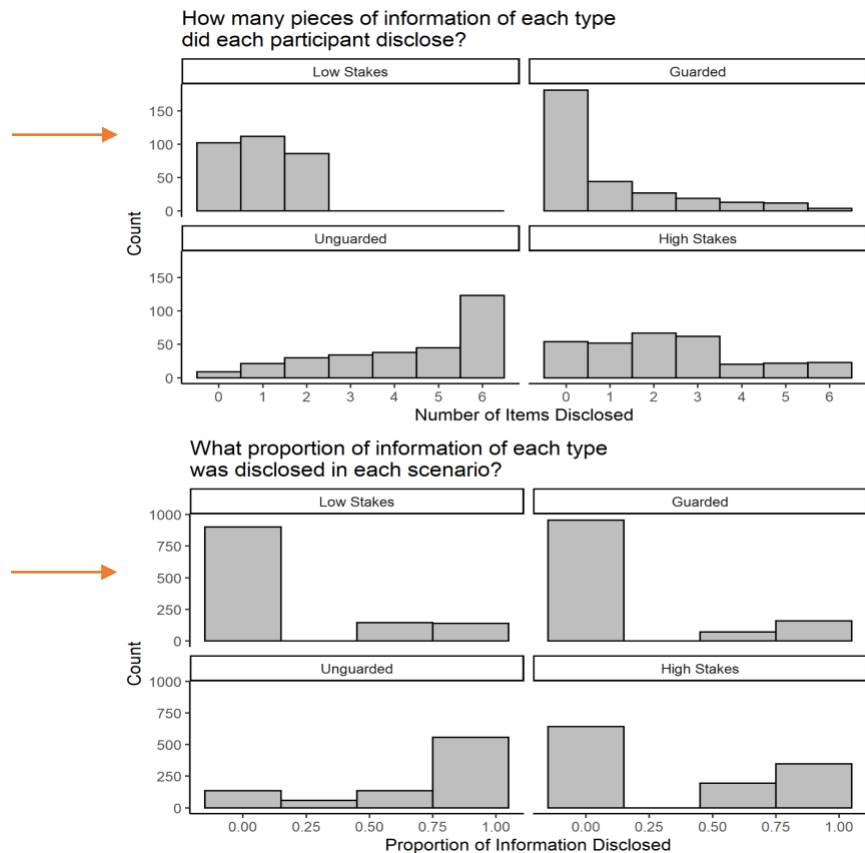
Many thanks for the constructive comments! In what follows, we clarify the issues raised.

Before getting into the specifics of the RR, you might find a recent experience of our using (another) online information management dilemma paradigm interesting (happy to discuss in a more informal context if useful). We ran a number of pilots where we presented a detailed/rich scenario (with varying 'seriousness' of information) and required participants to manage disclosure in a subsequent online interview. Our repeated pattern of findings across pilots and the actual study was that participants could be incredibly strategic about the kinds of information they were willing to 'give away' and what they were more reticent about providing. There was also a good deal of overlap in their strategies. So, based on our recent experience, I have two observations:

1. All illicit information is not equal.
   I like the way the DOM model formulates this – and I will take a closer look at the predictions when I have more time. However, our pilots showed that low stakes information was very 'off-loadable' in a context where the interviewees had to give something to avoid suspicion falling on themselves – in online contexts at least.
   We would be delighted to discuss your pilot because initially, we had also predicted that low-stakes information would be 'off-loadable'. However, the initial study on the DOM model revealed something different. Interviewees refrained from disclosing low-stakes information. That finding led us to refine the old prediction to the current one: people will refrain from disclosing low-stakes information because the corresponding benefit is low. So why take the risk, even if such risk is low, considering the low likelihood of a benefit. We believe that the tangible consequences of our research bring this property of low-stakes information to light. And it will be interesting to theorize in future work about how to shift the perception of low-stakes to unguarded. Below is a figure from our initial study of DOM.

How many pieces of information of each type did each participant disclose?

What proportion of information of each type was disclosed in each scenario?

2. Methodology matters.
   In a first pilot we use the forced-choice interview method you propose here – essentially to determine that our manipulation worked. I think it was useful to demonstrate that our information management dilemma actually produced a dilemma and resulted in reluctant interviewees. However, in a second pilot, we used open question responses. While we still found a pattern of reluctance (greater reluctance, if I recall correctly) there wasn't a like-for-like mapping. However, I entirely understand why you have opted for this format in an initial study – it makes sense to limit some of the wider strategies, at least in an initial test. I appreciate the exploration of the ecological validity issue in the RR but I agree entirely that as an initial step in formulating a methodology, it is important to insert control as necessary – and this can be gradually loosened in future work.

   Absolutely, we agree. Although, in the present design, participants can elect for a no comment response by selecting none of the information items. The plan for future work is to explore how we might manipulate costs and benefits in real-time while using verbal interviews.

Overall, with respect to the RR, I have the following comments:

1. "To what extent does network membership predict the type of information the interviewee will choose to reveal" – I don't think the design speaks to the specific way in which this question is formulated in the Introduction. The next sentence does clarify though ["Are two different people from the same network likely to disclose same information or does disclosure in this context better resemble individuals independently managing outcomes of revealing

information"]. I think sharpening up the Introduction will also help with reader understanding later (see additional comments below).

Thank you for drawing this issue to our attention. We have now edited the confusing phrase "network membership" out of the manuscript. And we have noted explicitly that we are referring to "the network a person belongs to". We appreciate that we might have misunderstood the comment here, and we are happy to revisit and clarify further if need be. Additionally, we have proofed the text to improve readability.

2. One thing I wasn't clear on pertains to the weighting of the information items (as % safe; % unsafe). Again, I completely understand the rationale for this approach having done some work on risk-taking/costs-benefits and elicitation recently. However, I got rather lost in the third para of p.12. If I've understood correctly, the 'weightings' of information will be randomly generated. Is this likely to be plausible/make sense to the participant? Are they likely to off-set some of this information and resort to decisions informed by plausibility based on general understanding of the world or too much Netflix (i.e. it is highly unlikely to matter in any way what X is, but Y is almost certainly going to be sensitive in some way)? Do you have any sense of this from piloting – if not, it might be worth piloting before burning resources on the whole study.

Yes, we have piloted the procedure, and it works pretty well (see Figure above, which is from Neequaye et al. 2021). Participants understand the instructions behave accordingly. However, the concern raised here is essential, which we considered to improve the present study's design. The initial study (i.e., Neequaye et al., 2021) randomized the cost and benefits across the board to induce the dilemma. We found that the narrative content substantially influenced people's decisions. So, plausibility based on a general understanding of the world matters. In the present design, we ensured that the narrative contents of the story also matched the plausibility of costs and benefits weightings. What the story suggests is dangerous or beneficial to disclose is also weighted as such at the decision-making stage.

3. Absent (largely) from the RR is any discussion of role/place in the network. While I appreciate adding in hierarchy is likely beyond the scope of this initial study, it is almost certainly the case that such networks produce hierarchies in which different members have different roles, responsibilities, and relationships. Such relationships are likely to play some role in how people make decisions about disclosure. Even in the experimental paradigm, the participant discussions will almost inevitable produce 'leaders' and 'followers'. I think it might be very difficult to build in at this stage but this issue should probably be highlighted within the background – even if only to offset later criticism. I like that you've included a measure of affiliation with the group - perhaps also asking some questions about such perceptions / leadership / hierarchy might be beneficial in terms of exploring and developing the methodology in the future.

We agree entirely that different features might characterize networks (e.g., hierarchy, roles, etc.). It is, indeed, challenging to manipulate all these possibilities at once in an initial study. Thus, we begin with or generic form of a network. A group consisting of a flat hierarchy wherein the members share a common illicit goal (This aspect is now noted explicitly on page 4). The affiliation measure will make for an interesting exploratory analysis and future design. As suggested, we will now include another exploratory measure asking whether any roles were created during the planning stages (see appendix).

4. Regarding the methodology in Phase 1, has money laundering actually taken place? My subsequent reading of the actual materials suggest it has (or at least there's been some dodgy activity) but the actual status of illicit behaviour should be clear in the Procedure also. In

general, I would suggest adding more detail throughout the Procedure (e.g., which three topics will they be interviewed about – and then provide the example, p.9) as clearly as possible. A Figure illustrating the procedure may also be helpful (especially for future replication).

Many thanks for flagging the issues here. We have now made explicit revisions to stimuli to ensure participants are aware that money laundering has taken place and is part of the group's modus operandi. All participants will be interviewed on all the three topics (presented in random order). This aspect is now noted explicitly in the procedure; sorry for the confusion. Participants will receive the story on all three topics during the Phase 1 planning and during the Phase 2 information disclosure stage. We have now included a figure on page 15 to better clarify the procedure. For conciseness, the figure outlines the highlights of the procedure. We are happy to include a qsf file with the submission to allow verbatim access to the more complex Phase 2 protocol.

5. With respect to the planning phase, will the group 'commit' somehow to what the plan is? If not, is it a risk to rely on individual participants later reporting it accurately (what if they misunderstood it or forgot it?)? Group discussions are notoriously difficult to parse for meaning/final outcome without some kind of explicit commitment phase (i.e., everyone has to agree to X or similar).

This concern is a thorny issue, indeed, and we hope to capture strategies via the open-ended responses at the end of the disclosure phase (i.e., Phase 2). We are concerned that an explicit commit phase after group planning might introduce a confound wherein participants stick to the group plan when they want to adjust their strategy. In the real world, network members might not necessarily commit to what their group planned. If such variation exists, we want to capture it in this research. A commitment phase reads to us more like a manipulation of LOYALTY, which is certainly interesting for future research. We will keep that in mind for future work, but we are happy to revisit the issue if we have missed anything.

6. Performance feedback (p.14) – when I first read this, I thought it meant that the individual interviewee would receive feedback about the on-going group performance average. Is that the case? If yes, I cannot see the rationale as it will likely confound the very phenomenon you want to look at. So, I'll assume this is individual level feedback with reference to the contribution to the group? This section needs some clarification.

The individual level feedback is with reference to a member's contribution to the group. We have now edited the confusing sentence on page 14.

7. The sample size appears well justified and the analysis plan looks broadly appropriate for a nested design. How/will you add any measures of affiliation or similar to this plan?

The affiliation measure is exploratory. We will run exploratory models including the first affiliation measure. We will explore the main effects as well as the interaction effects with the cost-benefit manipulations.

8. The writing style is over-complex / convoluted in places. Given the complexity of the design and procedure, a revision should work to be as precise and concise as possible. Similarly, I think the Introduction/Background might be clearer.

Many thanks; we have now tried to make the introduction more concise and include a figure summarizing the design. As noted earlier, we will now include a Qualtrics "qsf" file to allow verbatim access to the procedure in future replications.

References

Dawson, E. C. (2015). Improving Investigative Interviews: Facilitating Disclosure of Information through Implicit Means [Graduate Center, City University of New York]. https://academicworks.cuny.edu/gc_etds/900

Luke, T. J. (2021). A meta-analytic review of experimental tests of the interrogation technique of Hanns Joachim Scharff. Applied Cognitive Psychology, acp.3771. https://doi.org/10.1002/acp.3771

Neequaye, D. A. (2018). Eliciting information in intelligence interviews through priming: An examination of underlying mechanisms [Department of Psychology, University of Gothenburg]. https://gupea.ub.gu.se/handle/2077/57528

Neequaye, D. A., Luke, T. J., & Kollback, K. (2021). Managing Disclosure Outcomes in Intelligence Interviews. PsyArXiv. https://doi.org/10.31234/osf.io/tfp2c

Oleszkiewicz, S. (2016). Eliciting human intelligence a conceptualization and empirical testing of the Scharff-technique [Department of Psychology, University of Gothenburg]. http://hdl.handle.net/2077/41567

Soufan, A. (2011). The Black Banners: The Inside Story of 9/11 and the War Against al-Qaeda. W. W. Norton & Company.

Toliver, R. F. (1997). The interrogator: The story of Hans-Joachim Scharff, master interrogator of the Luftwaffe. Schiffer Pub.

Wertenbroch, K., & Skiera, B. (2002). Measuring Consumers' Willingness to Pay at the Point of Purchase. Journal of Marketing Research, 39(2), 228–241. https://doi.org/10.1509/jmkr.39.2.228.19086