**How Interviewees Determine What Interviewers Want to Know**

The appendices appear after the reference list.

**Author Note**

David A. Neequaye, Department of Psychology, Lancaster University, United Kingdom.

Alexandra Lorson, Center for Language and Cognition (CLCG), University of Groningen, Netherlands

Correspondence to: David A. Neequaye, Department of Psychology, Lancaster University, Lancaster LA1 4YF, United Kingdom. Email: [daneequaye@gmail.com](mailto:daneequaye@gmail.com)

**Data Availability Statement:** All data supporting the findings in this research will be publicly available here: <https://osf.io/jb3y6/>

There are no conflicts of interest to declare.

**Abstract**

We examine the mechanisms by which interviewees in investigative interviews mentally organize information when deciphering what an interviewer wants to know. The overarching idea is that such a process stems from the extent to which an interviewer’s question specifies an objective. Our initial test (i.e., Neequaye & Lorson, 2023) suggested two competing mechanisms: 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 propose two studies to ascertain which mechanism better captures how interviewees determine what interviewers want to know.

*Keywords*: disclosure; investigative interviewing; pragmatic correspondence; programmatic research

**How Interviewees Determine What Interviewers Want to Know**

Investigative interviews are formal social interactions wherein interviewers solicit information from interviewees in service of various issues (e.g., Neequaye, 2023). Such concerns include enhancing eyewitnesses’ ability to recall information, detecting lies, intelligence gathering, and insurance claims investigations (e.g., Fisher & Geiselman, 1992; Granhag & Hartwig, 2015; Hartwig et al., 2014; Warren & Schweitzer, 2018). Those domains of research concentrate on approaches interviewers use to elicit information. A fundamental assumption of this focus is that interviewees hypothesize about what their interviewers want to know. 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. For example, one of the primary goals of the Scharff interviewing technique is to conceal the interviewer’s objectives so that it is not apparent to interviewees that they are contributing to those objectives (see, e.g., Oleszkiewicz, 2016). Conversely, the Cognitive Interview encourages interviewers to make their objectives explicit and use mnemonic devices to enhance interviewees’ recall of accurate information (Fisher & Geiselman, 1992). *But how do interviewees discern what their interviewers want to know?* This vital step in the investigative interviewing process has received little attention. The literature needs an explanation of the mechanisms by which interviewees flag their interviewer’s objectives. Then, researchers can better specify how various interviewing methods might exert their effects.

Theorists in Pragmatics have theorized that when people converse, they piece together each other’s utterances to decode the messages being conveyed (e.g., Grice, 1989; McHoul, 1987). People pay more attention to utterances they believe will contribute a worthwhile difference to understanding the message being conveyed, and such perceived worthwhile utterances play a greater role in determining individuals’ comprehension of messages (Sperber & Wilson, 1987, 1995). And following Gricean maxims, people respond with messages that are relevant to the perceived subject(s) of discussion (Grice, 1975).

Drawing on the Pragmatics literature (i.e., Grice, 1975; McHoul, 1987; Sperber & Wilson, 1995), we proposed that interviewees determine their interviewer’s objectives based on the extent to which an interviewer’s question specifies an objective (see Neequaye & Lorson, 2023, for contextual details and the basis of the initial theory). We contended that high-specificity questions that clearly indicate what an interviewer wants to know would lead interviewees to *mentally* flag information items that pragmatically correspond to those objectives. By *pragmatically correspond*, we mean information items that, *if uttered*, will objectively provide the information an interviewer’s question requests—not necessarily more or less information. Low-specificity questions wherein an interviewer’s objectives appear broad will introduce more uncertainty regarding what the interviewer wants to know. By asking high- or low-specificity questions[[1]](#footnote-2), interviewers can influence the extent to which interviewees might discern what they want to know. In our initial proposal (i.e., Neequaye & Lorson, 2023), we limited the present theory to intelligence gathering. Here, we abandon that restriction and propose that the mechanisms apply to investigative interviews in general.

The following illustration provides an overview of the theory (see Figure 1). Imagine that multiple interviewees held the same body of information on a topic. Our contention is that the way an interviewer frames a question could lead those respective interviewees to focus on different aspects of that information corpus. High-specificity questions influence interviewees to focus on *particular* details (i.e., pragmatic correspondence) such that they ignore information that does not provide *the specific answer* a question requests. Conversely, low-specificity questions will make interviewees focus on a broader range of information items, given the uncertainty of the interviewer’s objectives. That is to say, 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. 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. The possibility of making interviewees witting or unwitting cooperators or resisters has crucial ethical and efficacy implications. We will revisit that issue later in the Discussion Section after our proposal has undergone the necessary testing.

A diagram of a drug

Description automatically generated

It is worth highlighting the distinction between the current research and related work, like memory activation and reporting (e.g., Koriat & Goldsmith, 1996). 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 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 encountered the bomb (memory issue). An unfortunate interaction between pragmatics and memory is worth mentioning here. It is possible for an interviewer to assume that their objective is clear, and an interviewee might confirm the interviewer's assumption. However, that confirmation could be mistaken for reluctance or uncooperativeness if the interviewee cannot answer the interviewer's question due to memory issues (e.g., encoding failure or forgetfulness)[[2]](#footnote-3). This possible conflict between pragmatic assumptions and memory highlights the need to better understand how interviewees determine what interviewers want to know. Then, researchers can advise practitioners that an interviewee acknowledging their information objectives does not automatically imply that the interviewee possesses the requisite information.

**Recapping an Initial Examination of the Theory: Neequaye and Lorson (2023)**

Initial tests suggested that the process by which interviewees hone in on information items might differ from the original theory just described (i.e., Figure 1). We (i.e., Neequaye & Lorson, 2023) found that interviewees assume their interviewer wants to know *all* the information they hold on a subject under discussion. Most interviewees indicated that their interviewer wanted to know everything regardless of whether the interviewer posed a high- or low-specificity question (Neequaye & Lorson, 2023; henceforth NL). However, interviewees were more confident that they had identified what their interviewer wanted to know when the interviewer posed high- versus low-specificity questions. We believe our previous procedure (i.e., NL, Study 1) raises issues worth addressing.

NL invited participants to act as informants in a criminal investigation. The relevant experiment (Study 1) comprised ten scenarios where participants discovered information about a gang under investigation. In each scenario, an interviewer posed a question to elicit information about the respective discoveries. The research employed a within-subjects design during the question-specificity trials. High-specificity questions always asked for the *complete details* participants held, while low-specificity questions were broad because they requested to know *any detail* about a discovery. 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. The mixture of high- and low-specificity questions plus the use of a predefined response list might have introduced a potential confound.

All things being equal, any investigator would want the complete details an interviewee holds, given that complete information would be more beneficial to any investigation than partial details. By mixing question-specificity (i.e., within-subject trials), it is possible that participants disregarded the uncertain nature of low-specificity questions. They might have inferred that their interviewer always wanted to know complete details because that same interviewer also frequently asked for complete details—the de facto purpose of any investigative interview. Moreover, NL’s design could have made such an assumption even more salient by presenting participants with a predefined list including complete details as a choice option. Taken together, NL’s design may have influenced participants to *always* hone in on complete details.

A further limitation of NL’s procedure (Study 1) is that pragmatic correspondence was designed to be equivalent to complete details such that complete details pragmatically corresponded to high-specificity questions. Put differently, high-specificity questions always asked for complete details in any given scenario (NL, Study 1). However, pragmatic correspondence does not necessarily denote complete or partial information. Pragmatic correspondence refers to *specific information* that, all things being equal, will truthfully answer an interviewer’s question; suppose the interviewee holds such information. As noted, NL’s protocol might have influenced interviewees to assume their interviewer always wanted to know complete details, regardless of question specificity. And because the experiment also equalized the status of pragmatic correspondence and complete details, the study may have failed to test whether high- versus low-specificity questions, indeed, draw attention to specific details.

**The Present Research**

The argument that NL’s procedure influenced participants to always hone in on complete details is a speculation that remains to be verified. Here, we reexamine NL’s original theory while addressing the contentions we raised about their procedure. The present proposal includes two replications of NL (Study 1). The findings will assist us in determining further appropriate studies to uncover the mechanisms by which interviewees determine what interviewers want to know.

Similar to NL (Study 1), the protocol of Replications 1 and 2 will equalize the status of pragmatic correspondence and complete details. But Replication 1 will implement a within-subjects design for question-specificity trials, and Replication 2 will employ a between-subjects design. Furthermore, in both replications, participants will indicate what their interviewer wants to know using a free text response rather than choosing from a predefined list.

Suppose Replications 1 and 2 generate NL’s (Study1) finding that interviewees focus on complete details, regardless of question-specificity. Then, NL’s original theory requires a significant revision. Such results would suggest that question-specificity has little influence on determining what an interviewer wants to know; interviewees likely assume that interviewers always want to elicit complete details. If Replication 1 replicates NL (Study1) but Replication 2 fails to replicate. Then, it is likely that the mixing of high- and low-specificity questions makes interviewees focus on complete details, which is not necessarily a de facto assumption that the purpose of any interview is to elicit complete details.

**Overview of Research Protocols and Hypotheses**

In both studies, participants will begin by assuming the role of an informant who an investigator has propositioned to inform on a drug-dealing gang. 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. The informants will make several discoveries about the gang’s operations, and the investigator will ask them about those discoveries. Then, the informant role becomes an interviewee role, and the investigator becomes an interviewer. Some of the interviewer’s questions will indicate an explicit objective (high-specificity), and the aim of others will be comparatively uncertain (low-specificity). Interviewees will *not* give direct answers to the questions.The instructions invite them to (1) indicate what they think their interviewer wants to know and (2) give a confidence rating on that choice. The overarching goal of the research design is to ascertain whether high- versus low-specificity questions influence interviewees to focus on specific information (i.e., pragmatic correspondence). Suppose NL’s original proposal has significant verisimilitude; the subsequent predictions should receive support.

When interviewees determine what an interviewer wants to know, high-versus low-specificity questions should elicit more designations of information items aligning with pragmatic correspondence (Core Hypothesis 1). That is, high- versus low-specificity questions should significantly influence the perceived specificity of interviewees’ responses. Low-versus high-specificity questions should make interviewees less confident that they have identified what their interviewer wants to know (Core Hypothesis 2). This confidence assessment is a proxy to examine whether low- versus high-specificity questions make it challenging to determine what an interviewer wants to know. Disposition should have no effect on what interviewees think their interviewer wants to know (Core Hypothesis 3).

NL’s original theory should be revised if the following pattern of results emerges rather than support for the Core Hypotheses. Question-specificity has no effect when interviewees determine what an interviewer wants to know (Revision Hypothesis 1). Question-specificity has no effect on interviewees’ confidence that they have flagged what their interviewer wants to know (Revision Hypothesis 2). Such findings will support the idea that interviewees assume their interviewer always wants to know all the information they possess. This potential revision to NL’s theory is a key aspect of the present research.

**Method: Replications 1 and 2**

**Sampling Plan and Power Analysis**

We will recruit a minimum of 600 English-speaking participants (*N* = 300, per experiment) via the Prolific Academic participant pool. We conducted simulations to examine the level of precision the chosen sample size for each Replication Study can provide, given our resources and planned hypotheses tests. The simulations indicated that the sample size will suffice (see the Analysis Plan).

Each experiment is expected to last approximately 15 minutes per participant (estimated compensation = £9/hr).

**Procedure**

We describe the protocols of the replications at once for the sake of conciseness. When necessary, we highlight the differences between the protocols. The project has received ethics approval (FST-2023-4117-RECR-4) at BLINDED. The replications will be conducted simultaneously to ensure that prospective participants do not partake in more than one experiment. The research will be entirely online, on Qualtrics, and introduced as studies about communication within a law enforcement context.

The procedure protocol can be reproduced using the Qualtrics (qsf) file available here <https://osf.io/qt4p3/>

**Phase 1: Informant Role.** Participants will read a background story to assume the role of an informant who can gather information about a drug-dealing gang. The plot is such that informants contend with the possibility of disclosing information to an interviewer, which is typical in investigative interviewing research (e.g., Oleszkiewicz, 2016). The story manipulates informants’ dispositions by inviting them to take on either a *cooperative*, *semi-cooperative*, or *resistant* mindset when engaging with their interviewer (see Appendix A).

**Phase 2: Introduction to Decision-making Instructions.** Next, participants will undergo an instruction stage to get acquainted with how to engage with the interviewer’s questions. An instructional manipulation check (IMC) will be included to identify and exclude inattentive participants who fail the check (see Appendix B). The instructions will tell participants that they will undergo a number of scenarios. In each scenario, they will discover information about the criminal gang being investigated. Then, they will receive a question from their interviewer about the earlier discovery. Participants will be told to write what they think their interviewer *wants to know*—not what they intend to disclose.

**Phase 3: Discoveries, Questions, and Decision-making**. The format and number of scenarios for the question-specificity trials will depend on the respective replications. Appendix C outlines the stimulus material in detail.

***Replication 1.*** Participants will undergo six scenarios in random order. In each scenario, they will discover something about the gang under investigation. Similar to NL, the discoveries will be such that participants can describe them in three legitimate ways: (*i*) bare minimum details; (*ii*) medium details—i.e., a new detail plus the bare minimum; or (*iii*) complete details—i.e., new detail plus the bare minimum and medium details. Consider this discovery: an informant overheard a phone call in which a gang member told a colleague, “It is better to sell the off-brand green-star oxycodone.” A substantive description of that discovery could embody any of the following contents: they sell *oxycodone* (bare minimum), *green-star oxycodone* (medium), or *an off-brand green-star oxycodone* (complete).

Replication 1 will employ a within-subjects design for the question-specificity trials. After each discovery, the interviewer will ask a high- or low-specificity question, three questions per condition—participants will undergo six trials. Following NL’s design, high-specificity questions will *specifically* request the complete details of a discovery. Low-specificity questions will ask for *anything* such that participants can reasonably think the interviewer wanted to know the bare minimum, medium, or complete discovery.

Next, participants will be invited to write what they think their interviewer wants to know. Participants will receive a summary of their discovery—to ensure that they do not forget the contents, given that the present research is not about memory. The summary will be followed by a textbox structured in a format to keep participants cognizant of the instruction to write what they think their interviewer’s objective is. The textbox will be preceded by the prompt “The police-contact wants to know if…” (see Figure 2). We settled on this protocol after conducting two pilot tests to determine how best to comprehensively capture participants’ thinking (see <https://osf.io/rvmn6/>).

A screenshot of a questionnaire

Description automatically generated

After participants indicate what they think the interviewer wants to know, they will provide two confidence ratings regarding their choice. One rating will be mandatory: On a scale from 1 – 5, how confident are you that “*participants verbatim text*” is what the police-contact wants to know (1 = *not confident at all*, 5 = *completely confident*)? The optional rating will invite participants to place a hypothetical wager on their choice being what the interviewer wants to know. The wager will be a percentage of their compensation (0% = *none of my compensation*, 100% = *all of my compensation*). 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.

***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. Participants will be randomly assigned to the high- or low-specificity condition.

***Exclusion criteria***

Both studies will include four control questions to flag the data of inattentive participants (see Appendix D). The control questions were randomly distributed among the scenarios in each study. We will exclude the data of participants who fail the IMC (see Appendix B) and those who fail one control question, respectively.

**Coding Strategy**

Two assistants, blind to the hypotheses, will code participants’ responses. The coders will receive a data file containing anonymized participant IDs along with their responses and the corresponding scenarios.

The assistants will use that data file to execute their codings in a separate data entry form developed with REDCap (Research Electronic Data Capture), a secure web-based software platform hosted at [BLINDED]. The coders will first enter a participant’s ID and indicate the specific scenario being coded. The scenario selection will reveal the contents of that scenario; then, the coders will input, from the data file, the participant’s response regarding the scenario. This extra step is to ensure that the coders remain aware of the response undergoing coding. Next, they will rate the extent to which the high- or low-specificity would better elicit (or better fit) the response being coded. This rating will be provided using a visual analog slider ranging from -100 to +100, including a zero (0) midpoint. The leftmost hand of the scale (i.e., -100) will display the low-specificity question, and the rightmost hand (i.e., +100) will display the high-specificity question. In this way, moving the slider to the left or right will code whether a participant’s response pragmatically corresponds more to the high- or low-specificity question. The descriptive label of the zero (0) mid-point will be “cannot decide”, and it will be used in the following situations: (1) when the response does not describe the contents of the scenario; (2) when the response says “no” or “I do not know”; or (3) “when the response fits neither the high- or low-specificity question”. Figure 3 depicts the coding, and the code book (i.e., data dictionary) can be accessed here: <https://osf.io/krybn/>

A screenshot of a computer

Description automatically generated

**Interrater Reliability**

Each coder will code 30% of the data for an interrater reliability assessment before testing the hypotheses. To examine the consistency between the coders’ ratings, we will calculate the intraclass correlation coefficient (ICC). The minimum ICC value to be accepted is 0.6; if it is below this value, the coders will discuss the causes of disagreements and resolve them independently. Then, they will recode the initial data until they achieve the minimum threshold, after which each coder will code 70% of the remaining data. When they complete their ratings, we will create an index of pragmatic correspondence using the average rating of both coders.

**Analysis Plan**

**Overview**

All the analyses will be conducted with the R software environment (The R Foundation, n.d.). We will test our hypotheses (i.e., pragmatic correspondence preference and confidence) using Bayesian linear regression models via the brms package version 2.21 (Bürkner, 2017), which provides an interface to fit Bayesian mixed models via Stan (Stan Development Team, 2017). Suppose our models prove unsuitable, given the data; we will employ other analysis approaches, such as the beta regression model. This contingency will be determined using posterior predictive simulations (i.e., the brms function pp\_check(), with ndraws = 100). Importantly, our decision regarding the data-generating process we assume to underlie the data will *not* be influenced by the hypothesis[[3]](#footnote-4).

Each analysis will produce posterior distributions over parameters, quantifying the probability of each possible parameter value, given the data. We will report the posterior mean with the corresponding 95% credible interval (95%-CrI) and the 95% highest density interval (HDI). The 95%-CrI is the range around the posterior mean within which the true value of the parameter lies with a probability of 0.95. The HDI is identical to the CrI if the posterior is symmetric; if the posterior is asymmetric, the endpoints of both intervals may differ.

Following best practices (Kruschke, 2014; Kruschke & Liddell, 2018; Vasishth et al., 2018), we defined a region of practical equivalence (ROPE) to examine whether the evidence that emerges from the findings is consistent with our predictions. The ROPE can be understood as a null region or a region encompassing parameter values that correspond to “no effect (given our theoretical propositions)”: effect sizes too small to be considered as supporting the hypotheses. We will assume an effect warrants consideration for our theoretical propositions if the corresponding parameter’s HDI falls outside the null region. In other words, if the 95% HDI falls outside the ROPE, it means that the 95% most credible values of that parameter are not practically equivalent to the null region (Kruschke, 2017). If a parameter’s HDI overlaps with the null region and the sign is positive, we will reject a theory postulating a negative effect. We will reject a theory postulating a positive effect if the sign is negative. If the parameter’s HDI falls entirely within the null region, we will conclude that the data are consistent with ‘no effect’ (not to say that we have proved that the null hypothesis is true). That instance means that the 95% most credible values of the parameter are practically equivalent to the null region (Kruschke, 2017). We will not settle on a conclusion from our data when the ROPE lies entirely within the parameter’s HDI.

Our resources will allow a sample size of 600 participants (*N* = 300, per experiment). Given that constraint and our planned hypotheses tests, we conducted simulations to examine the level of precision the chosen sample size can provide. Our desired level of precision is that the width of parameter coefficients’ 95% HDIs should be equal to or smaller than 16 (based on Neequaye & Lorson, 2023). We ran four models with simulated data, which indicated that a model with 270 participants could reach our desired precision: <https://osf.io/s5q7m/>

**Model Specification: Pragmatic Correspondence**

***Fixed and Random Effects***

To predict pragmatic correspondence preference, we will fit three truncated (lower bound -100 and upper bound = 100) Bayesian linear regression models, one per experiment (i.e., Models 1a and 2a) and one meta-analysis model (i.e., Model 3). The modes will be truncated to account for the specific range of our dependent variable (i.e., [-100, 100]) since predictions outside this range are meaningless. Models 1a and 2a will include the same fixed effects but different random effects structures. The two models will include the variables disposition (cooperative vs. semi-cooperative vs. resistant) and question type (high- vs. low-specificity) as predictors. We will add the interaction of both predictors for exploratory purposes. To examine Core Hypothesis 1/Revision Hypothesis 1a—for both models—the predictors will be sum-coded (question type: high-specificity = *1*, low-specificity = *-1;* disposition: cooperative = *1, 0*; resistant = *0, 1*, semi-cooperative = *-1, -1*). To test Core Hypothesis 3, the disposition variable will be treatment-coded with the cooperative condition as the reference level. Model 1a (Replication 1) will include random by-item slopes for question type and disposition, together with random by-participant slopes for question type. Conversely, Model 2a (Replication 2) will not include random by-participant slopes for question type since Replication 2 employs a between-subjects manipulation for question type.

Model 3 targets Revision Hypothesis 1b and tests whether the manipulation of question type—either as a within- or between-subjects factor—indeed influences the participants’ responses. To examine the effect of design type, the model includes the predictors design type, question type, and disposition, plus the interaction of design type and question type. The model will include random by-participant intercepts, and random by-items slopes for design type, question type, disposition, and the interaction of design type and question type.

***Priors***

For all models, we will use the same weakly regularizing priors, allowing a reasonably wide range of parameter values. The priors for the intercept will be normal distributions with mean 0 and standard deviation 20 based on the assumption that averaged over question type (and for semi-cooperative participants), the perceived specificity should be centered around zero. Skepticism in this assumption is introduced by defining a relatively broad prior of 30. For fixed effects, normal priors with a mean of 0 and a standard deviation of 20 will be used. This prior is conservative because it assigns most probability mass to values close to zero, which would correspond to a null effect. Such an approach demonstrates our commitment to giving the Revision Hypothesis a worthwhile chance of receiving support. Random effects will be modeled as a correlation matrix and a vector of standard deviations. The standard deviations will be assigned half-normal priors with a mean of 0 and a standard deviation of 1. For the correlation matrix, an LKJ(2) prior will be used such that smaller correlations are favored over extreme values such as ±1 (Stan Development Team, 2017; Sorensen et al., 2016). A prior-sensitivity analysis will be carried out to assess whether priors are dominating the posterior distribution.

***Region of Practical Equivalence and Model Comparison***

The specified ROPE of [-8, 8] is based on NL’s initial finding about the effect of low- versus high-specificity questions on participants’ decisions regarding what an interviewer wants to know (i.e., Neequaye & Lorson, 2023). NL’s results indicated that the lowest possible difference between the question type conditions was 4%. Applying that finding to the outcome variable of the present research (a continuous value ranging from -100 to 100), a 4% difference translates to a difference of 8 and a ROPE of [-8, 8]. We will specify the same ROPE for the effect of disposition and the interaction terms. If the data does not fit with a normal distribution and calls for a different model, we will consider a ROPE of [-0.17, 0.17] on the log-odds scale directly derived from Neequaye & Lorson (2023).

***Sampling Process***

We will draw samples from the posterior distributions of the model parameters using the NUTS sampler (Hoffman & Gelman, 2014). Four sampling chains will be run, each collecting 4,000 iterations. The first 1,000 iterations will be disregarded as part of the warm-up phase, leading to 12,000 iterations available for analysis. This sampling process should be the same for all models, and the chains should mix well (all R = 1.0).

In case our models show divergent transitions after warm-up, we will follow suggested solutions to resolve divergences that involve changing the MCMC criteria, for example, by raising adapt\_delta, increasing the number of iterations, or increasing tree depth, etc. Suppose we run into convergence issues and this model formulation turns out to be impossible to estimate without divergent transitions even after tweaking the MCMC criteria; we will assess whether dropping by-item or by-participant random slopes achieves convergence.

***Predictions***

To support Core Hypothesis 1, we expect the following result. High-specificity questions should elicit a greater preference for pragmatic correspondence than low-specificity questions. The 95% HDI of the test parameter should fall completely outside the ROPE, and considering that the question type variable will be sum-coded (high-specificity = *1*, low-specificity = *-1*), the test parameter’s 95% HDI should have a positive sign. If the parameter’s 95% HDI falls completely within the specified ROPE, we will interpret our findings as being consistent with Revision Hypothesis 1a. Furthermore, we will consider our findings consistent with Revision Hypothesis 1b, if, for Model 3, the interaction parameter’s 95% HDI falls completely inside the specified ROPE. We will count this finding as evidence against the claim that high-specificity questions only elicit greater pragmatic correspondence if the interviewer consistently uses high-specificity questions (as mirrored by Replication 2’s between-subjects design).

According to Core Hypothesis 3, the disposition variable should have no effect on preference for pragmatic correspondence when determining what an interviewer wants to know. The 95% HDI of the test parameter should fall completely within the ROPE, such that the 95% most credible values are practically equivalent to what we consider to be negligible effect sizes.

**Model Specification: Confidence**

***Fixed and random effects***

We will examine confidence ratings by fitting two mixed-effects Bayesian ordinal (cumulative) regression models (Models 1b and 2b). Question type and disposition will be included as predictors. Model 1b (i.e., Replication 1) will include varying intercepts and slopes for participants and items, assuming that the effect of question type on confidence ratings varies between participants and scenarios. Model 2b (i.e., Replication 2) will only include varying intercepts for participants as both disposition and question type will be between-subjects factors.

***Priors***

We will use the default priors of brms except for the priors of the intercept and main condition effect. We will use weakly regularizing priors, which will allow a reasonably wide range of parameter values while penalizing very extreme values. The prior for the intercept will be normally distributed with mean 0 and standard deviation 1. Furthermore, to build mild skepticism into our models, we will set a weakly informative prior on the condition effect (Lemoine, 2019; McElreath, 2020): a normal distribution centered at zero with a standard deviation of 0.5. The LKJ prior is set to 2. A prior-sensitivity analysis will be carried out to assess whether priors are dominating the posterior distribution.

***Prediction***

Core Hypothesis 2 will receive support if low- versus high-specificity questions lead interviewees to be less confident that they have flagged what their interviewer wants to know—that difference should emerge, regardless of disposition. The test parameter’s 95% HDI should fall completely outside the ROPE, and because the question type variable will be sum-coded (high-specificity = *1*, low-specificity = *-1*), the parameter’s 95% HDI should have a positive sign. . If the parameter’s 95% HDI falls completely within the specified ROPE, we interpret our findings as being consistent with Revision Hypothesis 2.

**Model Specification: Wager (i.e., Secondary Confidence Variable)**

We included a second confidence measure to assess the willingness of participants to place a bet that their preference is what the interviewer wanted to know. Analogous to the confidence ratings, we predict that low- versus high-specificity questions will lead to a lower probability of betting, independent of disposition. We will report the posterior mean, the 95% credible interval (95%-CrI), and the probability that a given coefficient is greater than zero, given the data and model. This analysis is exploratory and will be reported as such.

***Fixed, random effects, and Priors***

We will examine participants’ willingness to place a wager by fitting two mixed-effects Bayesian logistic regression models (Models 1c and 2c). Question type and disposition will be included to predict wagers. Model 1c (i.e., Replication 1) will include varying intercepts and slopes for participants and items, assuming that the effect of question type on confidence ratings varies between participants and scenarios. Model 2c (i.e., Replication 2) will include varying intercepts for participants only because disposition and question type will be between-subjects factors.

We will employ the same prior structure as described for the primary confidence measure.

**Study Design Template**

**Notes:**

1. The outcome measure for Models 1a and 2a is a continuous variable ranging from -100 to 100.
2. **Sampling plan and test sensitivity rationale:** We aim to remain with a minimum of *N* = 600 participants, *N* = 300 per study. Resource constraints and previous research (to precisely estimate a Region of Practical Equivalence [ROPE]) determined our sample size choice. For Replications 1 and 2, it holds that we will not conclude anything from our data when the ROPE lies entirely within the parameter’s HDI.
3. **Theory that could be shown wrong by outcomes:** Our goal is to ascertain the verisimilitude of two competing mechanisms following Neequaye and Lorson’s findings (2023). *Mechanism-1*: 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—versus—*Mechanism-2*: Interviewees generally assume that interviewers want to know all the information at their disposal, irrespective of question specificity. We cannot rule out plausible effects that are smaller than the limits of our ROPE. Nonetheless, supporting the respective core hypotheses will count as support for Mechanism-1, but rejecting the core hypothesis will count as support for Mechanism-2.

**Table X1**

*Replication 1: Question Type as a within-subjects factor*

|  |  |  |  |
| --- | --- | --- | --- |
| **Hypothesis** | **Model 1a** | **Analysis** | **Predictions** |
| **Core hypothesis 1**  High-versus low-specificity questions should elicit more designations of information items that align with pragmatic correspondence | brm(Specificity | trunc(ub = 100, lb = -100) ∼ Disposition + QuType + (QuType | SubjectID) + (Disposition + QuType | Context)  Contrast coding for Model 1a:  **Question-type:**  high-specificity = 1,  low-specificity = -1  **Disposition:**  cooperative = 0 1  resistant. = 1 0  semi-coop. = -1 -1  The model for Replication 1 will only include the two predictors Question Type and Disposition and no interaction term. A model including an interaction term will be run for exploratory purposes. | To test this hypothesis, we will investigate whether there is a main effect of question-type on the perceived specificity of participants’ responses. | The Question Type parameter’s HDI should lie outside the ROPE and have a positive sign for high-specificity questions (which are coded as 1). |
| **Revision Hypothesis 1a**  High-versus low-specificity questions do not elicit more designations of information items that align with pragmatic correspondence. | brm(Specificity | trunc(ub = 100, lb = -100) ∼ Disposition + QuType + (QuType | SubjectID) + (Disposition + QuType | Context)  Contrast coding for Model 1a:  **Question-type:**  high-specificity = 1,  low-specificity = -1  **Disposition:**  cooperative = 0 1  resistant. = 1 0  semi-coop. = -1 -1 | To test this hypothesis, we will investigate whether there is a main effect of question-type on the perceived specificity of participants’ responses. | The Question Type parameter’s HDI is predicted to fall within the null region, such that we can conclude the data are consistent with ‘no effect’ of question-type (not to say that we have proven that the null hypothesis is true). |
| **Core hypothesis 3**  There should be no effect of disposition on preference for pragmatic correspondence. | brm(Specificity | trunc(ub = 100, lb = -100) ∼ Disposition + QuType + (QuType | SubjectID) + (Disposition + QuType | Context)  Contrast coding for Model 1a:  **Question-type:**  high-specificity = 1,  low-specificity = -1  **Disposition (treatment):**  cooperative = 0 0  resistant = 1 0  semi-coop. = 0 1 | To test this hypothesis, we will investigate whether there is a main effect of disposition on the perceived specificity of participants’ responses. | All the Disposition parameter's HDIs are predicted to fall within the null region, such that we can conclude the data are consistent with ‘no effect’ of disposition (not to say that we have proven that the null hypothesis is true). |

**Table X2**

*Replication 2: Question Type as a between-subjects factor*

|  |  |  |  |
| --- | --- | --- | --- |
| **Hypothesis** | **Model 2a** | **Analysis** | **Predictions** |
| **Core hypothesis 1**  High-versus low-specificity questions should elicit more designations of information items that align with pragmatic correspondence | brm(Specificity | trunc(ub = 100, lb = -100) ∼ Disposition + QuType + (1 | SubjectID) + (Disposition + QuType | Context)  Contrast coding for Model 2a:  **Question-type:**  high-specificity = 1,  low-specificity = -1  **Disposition:**  cooperative = 0 1  resistant. = 1 0  semi-coop. = -1 -1  The model for Study 2 will only include the two predictors Question Type and Disposition and no interaction term. A model including an interaction term will be run for exploratory purposes. | To test this hypothesis, we will investigate whether there is a main effect of question-type on the perceived specificity of participants’ responses. | The Question Type parameter’s HDI should lie outside the ROPE and have a positive sign for high-specificity questions (which are coded as 1). |
| **Revision Hypothesis 1a**  High-versus low-specificity questions do not elicit more designations of information items that align with pragmatic correspondence. | brm(Specificity | trunc(ub = 100, lb = -100) ∼ Disposition + QuType + (QuType | SubjectID) + (Disposition + QuType | Context)  Contrast coding for Model 1a:  **Question-type:**  high-specificity = 1,  low-specificity = -1  **Disposition:**  cooperative = 0 1  resistant. = 1 0  semi-coop. = -1 -1 | To test this hypothesis, we will investigate whether there is a main effect of question-type on the perceived specificity of participants’ responses. | The Question Type parameter’s HDI is predicted to fall within the null region, such that we can conclude the data are consistent with ‘no effect’ of question-type (not to say that we have proven that the null hypothesis is true). |
| **Core hypotheses 3**  There should be no effect of disposition on preference for pragmatic correspondence. | brm(Specificity | trunc(ub = 100, lb = -100) ∼ Disposition + QuType + (QuType | SubjectID) + (Disposition + QuType | Context)  Contrast coding for Model 1a:  **Question-type:**  high-specificity = 1,  low-specificity = -1  **Disposition (treatment):**  cooperative = 0 0  resistant = 1 0  semi-coop. = 0 1 | To test this hypothesis, we will investigate whether there is a main effect of disposition on the perceived specificity of participants’ responses. | All the Disposition parameter's HDIs are predicted to fall within the null region, such that we conclude that the data are consistent with ‘no effect’ of Disposition (not to say that we have proven that the null hypothesis is true). |

**Table X3**

*Investigating Question Type in interaction with Design Type*

|  |  |  |  |
| --- | --- | --- | --- |
| **Hypothesis** | **Model 1a** | **Analysis** | **Predictions** |
| **Revision Hypothesis 1b**  High- versus low-specificity questions manipulated as a between-subjects versus within-subjects factor do not elicit more designations of information items that align with pragmatic correspondence. | brm(Specificity | trunc(ub = 100, lb = -100) ∼ Disposition + QuType + DesignType + QuType:DesignType + (1 | SubjectID) + (Disposition + QuType + DesignType + QuType:DesignType | Context)  Contrast coding for Model 3:  **Question-type:**  high-specificity = 1,  low-specificity = -1  **Disposition:**  cooperative = 0 1  resistant. = 1 0  semi-coop. = -1 -1  **Design-type:**  between = 1  within = -1 | To test this hypothesis, we will investigate whether there is an interaction effect of question-type and design-type on the perceived specificity of participants’ responses. | The interaction parameter’s HDI is predicted to fall within the null region, such that we can conclude the data are consistent with ‘no effect’ of question type x design type (not to say that we have proven that the null hypothesis is true). |

**Confidence in decisions (Replication 1)**

**Table X3**

*Replication 1: Confidence Ratings*

Output (5-point Likert Scale): 'not confident at all', 'slightly confident', 'somewhat confident', 'fairly confident', 'completely confident'

|  |  |  |  |
| --- | --- | --- | --- |
| **Hypothesis** | **Model 1b (Ordinal cumulative model)** | **Analysis** | **Predictions** |
| **Core hypotheses 2**  High-versus low-specificity question should make participants more confident in their designation choices, independent of disposition. | brm(Confidence ∼ Disposition + QuType + (QuType | SubjectID) + (Disposition + QuType | Context)  Contrast coding for Model 1b:  **Question-type:**  high-specificity = 1,  low-specificity = -1  **Disposition (changes to treatment coding to assess disposition, see above):**  cooperative = 0 1  resistant. = 1 0  semi-coop. = -1 -1  The model included the two predictors Question Type and Disposition and no interaction term. A model including an interaction term was run for exploratory purposes. | To test this hypothesis, we will investigate whether there is a main effect of Question Type on the confidence ratings. | The Question Type parameter’s HDI should lie outside the ROPE and have a positive sign for high-specificity questions (which are coded as 1).  The Disposition parameter’s HDI is predicted to fall within the null region, such that we conclude that the data are consistent with “no effect” of Disposition on the confidence ratings (not to say that we have proven that the null hypothesis is true). |

**Table X4**

*Replication 1: Willingness to bet (exploratory)*

Output: willingness to bet, ‘yes’ (1), ‘no’ (0)

|  |  |  |  |
| --- | --- | --- | --- |
| **Hypothesis** | **Model 1c (Ordinal Cumulative model)** | **Analysis** | **Predictions** |
| High- versus low-specificity questions should increase the probability of betting, independent of disposition. | brm(Bet ∼ Disposition + QuType + (QuType | SubjectID) + (Disposition + QuType | Context)  Contrast coding for Model 1C:  **Question-type:**  high-specificity = 1,  low-specificity = -1  **Disposition (changes to treatment coding to assess disposition, see above):**  cooperative = 0 1  resistant. = 1 0  semi-coop. = -1 -1  The model included the two predictors Question Type and Disposition and no interaction term. A model including an interaction term was run for exploratory purposes. | To test this hypothesis, we will investigate whether there is a main effect of Question Type on the confidence ratings. | The Question Type parameter's HDI should lie outside the ROPE and should have a positive sign for high-specificity questions (which are coded as 1).  The Disposition parameter’s HDI is predicted to fall within the null region, such that we conclude that the data are consistent with “no effect” of Disposition on the confidence ratings (not to say that we have proven that the null hypothesis is true). |

**Confidence in decisions (Replication 2)**

**Table X5**

*Replication 2: Confidence Ratings*

Output (5-point Likert Scale): 'not confident at all', 'slightly confident', 'somewhat confident', 'fairly confident', 'completely confident'

|  |  |  |  |
| --- | --- | --- | --- |
| **Hypothesis** | **Model 2b (Ordinal cumulative model)** | **Analysis** | **Predictions** |
| **Core hypotheses 2**  High-versus low-specificity question should make participants more confident in their designation choices, independent of disposition. | brm(Confidence ∼ Disposition + QuType + (1| SubjectID) + (Disposition + QuType | Context)  Contrast coding for Model 2B:  **Question-type:**  high-specificity = 1,  low-specificity = -1  **Disposition (changes to treatment coding to assess disposition, see above):**  cooperative = 0 1  resistant. = 1 0  semi-coop. = -1 -1  The model included the two predictors Question Type and Disposition and no interaction term. A model including an interaction term was run for exploratory purposes. | To test this hypothesis, we investigated whether there was a main effect of question-type on the perceived specificity of the responses. | The Question Type parameter’s HDI should lie outside the ROPE and have a positive sign for high-specificity questions (which are coded as 1).  The Disposition parameter’s HDI is predicted to fall within the null region, such that we conclude that the data are consistent with “no effect” of Disposition on the confidence ratings (not to say that we have proven that the null hypothesis is true). |

**Table X6**

*Replication 2: Willingness to bet (exploratory)*

Output: willingness to bet, ‘yes’ (1), ‘no’ (0)

|  |  |  |  |
| --- | --- | --- | --- |
| **Hypothesis** | **Model 2c (Ordinal Cumulative model)** | **Analysis** | **Predictions** |
| High- versus low-specificity questions should increase the probability of betting, independent of disposition. | brm(Bet ∼ Disposition + QuType + (1 | SubjectID) + (Disposition + QuType | Context)  Contrast coding for Model 2C:  **Question-type:**  high-specificity = 1,  low-specificity = -1  **Disposition (changes to treatment coding to assess disposition, see above):**  cooperative = 0 1  resistant. = 1 0  semi-coop. = -1 -1  The model included the two predictors Question Type and Disposition and no interaction term. A model including an interaction term was run for exploratory purposes. | To test this hypothesis, we will investigate whether there is a main effect of Question Type on the participants’ willingness to bet. | The Question Type parameter's HDI should lie outside the ROPE and should have a positive sign for high-specificity questions (which are coded as 1).  The Disposition parameter’s HDI is predicted to fall within the null region, such that we conclude that the data are consistent with “no effect” of Disposition on the confidence ratings (not to say that we have proven that the null hypothesis is true). |

References

Bürkner, P.-C. (2017). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411. <https://arxiv.org/abs/1705.11123v2>

Fisher, R. P., & Geiselman, R. P. (1992). *Memory Enhancing Techniques for Investigative Interviewing. Springfield, Illinois: Charles C*. Charles C. Thomas.

Granhag, P. A., & Hartwig, M. (2015). The Strategic Use of Evidence Technique: A Conceptual Overview. In A. Vrij & B. Verschuere (Eds.), *Deception detection: Current challenges and new directions* (pp. 231–251). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118510001.ch10>

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.

Hartwig, M., Meissner, C. A., & Semel, M. D. (2014). Human Intelligence Interviewing and Interrogation: Assessing the Challenges of Developing an Ethical, Evidence-based Approach. In R. Bull (Ed.), *Investigative Interviewing* (pp. 209–228). Springer New York. <https://doi.org/10.1007/978-1-4614-9642-7_11>

Hoﬀman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–1623.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*(3), 490–517. <http://dx.doi.org.ezproxy.ub.gu.se/10.1037/0033-295X.103.3.490>

Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. <https://doi.org/10.3758/s13423-016-1221-4>

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912–928. <https://doi.org/10.1111/oik.05985>

McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>

Meissner, C. A., Kleinman, S. M., Mindthoff, A., Phillips, E. P., & Rothweiler, J. N. (2023). Investigative Interviewing: A Review of the Literature and a Model of Science-Based Practice. In D. DeMatteo & K. C. Scherr (Eds.), *The Oxford Handbook of Psychology and Law* (p. 0). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780197649138.013.34>

Neequaye, D. A. (2023). Why Rapport Seems Challenging to Define and What to Do About the Challenge. *Collabra: Psychology*, *9*(1), 90789. <https://doi.org/10.1525/collabra.90789>

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>

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>

Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology*, *12*(3), 175–200. <https://doi.org/10.20982/tqmp.12.3.p175>

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.

Stan Development Team. (2017). *Stan* [Computer software]. [//mc-stan.org/](https://doi.org/mc-stan.org/)

The R Foundation. (n.d.). *R: The R Project for Statistical Computing* [Computer software]. Retrieved May 31, 2024, from <https://www.r-project.org/>

Vasishth, S., Mertzen, D., Jäger, L. A., & Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language*, *103*, 151–175. <https://doi.org/10.1016/j.jml.2018.07.004>

Warren, D. E., & Schweitzer, M. E. (2018). When Lying Does Not Pay: How Experts Detect Insurance Fraud. *Journal of Business Ethics*, *150*(3), 711–726. <https://doi.org/10.1007/s10551-016-3124-8>

**Appendices**

**Introduction**

This study is about communication within a law enforcement context. You will read some fictional scenarios assuming the role of the main character. Then you will answer some questions about each scenario. **Most of this study involves reading. So, please read the scenarios and instructions carefully because understanding them is crucial**. We have included questions to check if you read and answered questions with your full attention.

The entire study will take approximately 10 - 15 minutes to complete. You will receive a compensation of £2.25 for participating once the study is over.

***Appendix A***

**Disposition manipulations**

Imagine that you are one of the owners of a restaurant in town; you also work at this restaurant, which overlooks a big park. You and your colleagues have a good picture of what goes on in the park. It is well known among the restaurant staff that a narcotics-dealing gang called KET22 operates in the park. Recently, a police-contact approached you and your colleagues to provide information about the gang if you discovered anything. The police-contact mentioned that none of you are obliged to give any information. <**Dispositional variations begin here**>

**Cooperative:** However, KET22 disrupts your business at the restaurant. So, it is in your best interest to assist the police in their investigations to eliminate the gang. Then, your business can grow.

**Semi-cooperative:** KET22 disrupts your business at the restaurant. Assisting the police in their investigations could eliminate the gang. Then, your business can grow. But KET22 might retaliate if they find out you are helping the police take them down. So, you must strike a good balance between assisting the police and safeguarding yourself.

**Resistant:** The police-contact does not know this, but because you (personally) came into some financial troubles, you occasionally supply narcotics to customers at the restaurant on the gang’s behalf. If the gang gets busted, you are very likely to get in trouble, too. You only agreed to meet with the police-contact to avoid suspicion.

**Manipulation check (disposition)**

Now you know your character or the role you are to play in this study. Suppose you were to make discoveries that could get KET22 busted, and the police-contact asked you about those discoveries. How would you engage with the interviewer?

* I will lie to ensure that I hide what I know. (-1)
* I will keep silent and not respond to the question. (0)
* I will reveal some of my discoveries, not everything I know. (1)
* I will reveal what I know. (2)

***Appendix B***

**Introduction to scenarios**

In the next phase of the study, you will be placed in various scenarios where you will make various discoveries about KET22, the gang under investigation. After each discovery, you will receive a question from the police contact about the discovery.

Your task in the upcoming phase is to **indicate** **what you think the police-contact WANTS TO KNOW about your discovery based on the police-contact’s question**. The task is NOT about indicating what you will necessarily say in response to the question.

**Your task is to indicate what you think the police-contact wants to know based on the police-contact’s question!**

We have included other questions to check if you read and answered the questions with your full attention.

**Instructional manipulation check**

What is **TRUE about your main task** in the upcoming phase?

* My task is to indicate what I think the police-contact wants to know based on the police-contact’s question. [**Pass]**
* My task is to indicate what I want to say in response to the police-contact’s question. **[Fail]**

Next follows the scenarios.

***Appendix C***

**Replication 1**

To facilitate randomization, participants will be randomly assigned to one of two lists. The lists’ contents will be presented in random order in both the utterance and designation conditions.

|  |  |  |
| --- | --- | --- |
|  | List 1 | List 2 |
| Scenario | 1a | 1b |
| Scenario | 2b | 2a |
| Scenario | 3a | 3b |
| Scenario | 4b | 4a |
| Scenario | 5a | 5b |
| Scenario | 6b | 6a |

**Scenarios - Replication 1**

a. High-specificity questions are highlighted in green.

b. Low-specificity questions are highlighted in yellow.

1. 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.”
   1. Have you discovered *the particular brand of narcotics* KET22 sells?
   2. Have you discovered anything about the gang’s narcotics sales lately?
2. 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.
   1. Do you know *the full details about the vehicle the KET22* gangsters usually arrive in at the park?
   2. Do you have any information about KET22’s transportation in the park?
3. 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.
   1. Have you spotted *the exact location* at the park where KET22 deals drugs?
   2. Have you spotted anything about where KET22 deals drugs?
4. On your way home after work, you saw that some KET22 gangsters were arguing. It was around 19.00 on Monday. From what you heard, the argument was about whether to sell a high dose of drugs to a VIP customer.
   1. Have you caught *the contents of particular interactions* between the gang members lately?
   2. Have there been any developments with the gang members lately?
5. At work last week, your colleague, who is becoming friends with a KET22 gangster, slipped you some details. She said that KET22 is connected to a much bigger gang called TETO. TETO supplies opioids wholesale.
   1. Do you have information about the *sources from which KET22 obtains narcotics*?
   2. Has anything about KET22’s narcotics operations come to your attention?
6. During one of your short breaks at work, you decided to enjoy some sunshine. So, you went to the edge of the park where there are benches. As you approached, you saw a rowdy group at one of the benches, and you chose the bench furthest away from them. The group was talking about how to contact KET22 to buy narcotics. They said customers could make contact by sending a text message containing a lion emoji to any KET22 phone number.
   1. Have you made observations about *exactly how customers contact* KET22 to buy narcotics?
   2. Have you made any observations about KET22’s customers?

**Replication 2**

This study will employ a between-subjects design.

**Scenarios - Replication 2**

a. High-specificity questions are highlighted in green.

b. Low-specificity questions are highlighted in yellow.

1. 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.”
   1. Have you discovered *the particular brand of narcotics* KET22 sells?
   2. Have you discovered anything about the gang’s narcotics sales lately?
2. 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.
   1. Do you know *the full details about the vehicle the KET22* gangsters usually arrive in at the park?
   2. Do you have any information about KET22’s transportation in the park?
3. 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.
   1. Have you spotted *the exact location* at the park where KET22 deals drugs?
   2. Have you spotted anything about where KET22 deals drugs?
4. On your way home after work, you saw that some KET22 gangsters were arguing. It was around 19.00 on Monday. From what you heard, the argument was about whether to sell a high dose of drugs to a VIP customer.
   1. Have you caught *the contents of particular interactions* between the gang members lately?
   2. Have there been any developments with the gang members lately?
5. At work last week, your colleague, who is becoming friends with a KET22 gangster, slipped you some details. She said that KET22 is connected to a much bigger gang called TETO. TETO supplies opioids wholesale.
   1. Do you have information about the *sources from which KET22 obtains narcotics*?
   2. Has anything about KET22’s narcotics operations come to your attention?

**Confidence rating (After each scenario)**

* The police-contact asked:

[*display question*]

* Based on the above question, **you wrote**: the police-contact wants to know if

[*display selection*]

* On a scale from 1 – 5, how confident are you that **what you wrote** the police-contact wants to know?

1 = *not confident at all*, 2 = *slightly confident*, 3 = *somewhat confident*, 4 = *fairly confident*, 5 = *completely confident*

**Confidence measure via bets [optional question]**

* Suppose you were to **place a bet on your text below**.
* **You wrote** the police contact wants to know if [*display text*]
* On a scale from 0 to 100, what percentage of your compensation (for participating in this research) are you willing to bet that **what you wrote** is what the police-contact wants to know.

0% = none of my compensation, 100% = all of my compensation

***Appendix D***

**Control Questions**

*The control questions will employ the same scenario outlined below. The scenario will be presented four times, in random order, with four different questions.*

Recently, a man came into the restaurant to buy coffee. You suspect he might be one of the KET22 gangsters, but you are unsure. When he made his order, there was no milk at the counter. So, you asked your colleague to get some milk from the fridge in the back. While you were waiting, you got a good look at his face and stature. You can guess that he is about 190cm tall. His hair was dark with grey streaks. He had green eyes and a scar on his left jaw. The name on the card he used to pay for his drink was Kari Jupo.

Q1. From the options below, select the name on the card the man used to pay for his drink.

* Minea Blankson
* Johnny Biles
* Kari Jupo
* Renave Olsson

Q2. From the options below, select the correct description of the man’s hair.

* Blonde with brown streaks
* Blonde with grey streaks
* Dark with grey streaks
* Dark with yellow streaks

Q3. From the options below, select what the man ordered.

* Sandwich
* Coffee
* Beer
* Salad

Q4. From the options below, select the correct description of the man’s height

* 190cm
* 200cm
* 164cm
* 175cm

1. In previous work, we used the phrases high-worthwhileness and low-worthwhileness questions. Now, we use the phrases high-specificity and low-specificity, as *specificity* is a more intuitive suffix than *worthwhileness*. [↑](#footnote-ref-2)
2. Many thanks to Feni Kontogianni for drawing our attention to this issue. [↑](#footnote-ref-3)
3. We will use another model if the predictive draws deviate from our data, indicating that a revised approach to testing the hypothesis is warranted. Such a decision cannot be determined objectively until the actual analysis: we might encounter a bimodal distribution, making it challenging to run our models under the assumption of a normal distribution. An example of being influenced by the hypothesis is that a normal distribution underlies our data. Then, we use linear regression, but after failing to support the hypothesis, we change to a beta model to fish for effects despite the linear regression model being the appropriate tool. We will *not* implement such questionable practices. [↑](#footnote-ref-4)