

Epistemicity and communicative strategies

Abstract

When communicating, interlocutors negotiate knowledge by proposing propositional content to be added to their shared common ground. The way in which speakers put forward propositional content – expressing more or less confidence in its truthfulness – may affect the way in which other interlocutors react to such content. This paper examines speakers' production choices and hearers' interpretations of the formulations *believe*/*know*/bare assertion to test how (maximal) certainty is expressed and inferred, whether speakers adjust their production choices depending on their communicative goals and whether hearers are able to adjust their interpretations correspondingly. For this purpose, we created two scenarios – one with a cooperative interlocutor and one with a potentially uncooperative interlocutor. The results suggest that *know* is epistemically the most powerful formulation – stronger than the bare assertion – but that the bare assertion may still be preferred over *know* for expressing maximal certainty in cooperative scenarios. Our findings also suggest that *believe* is used to hedge the assertive strength of statements in cooperative settings. Whereas speakers and hearers agree in the relative epistemic ordering of the formulations (*believe*<bare assertion<*know*), when inferring the speakers' degrees of belief hearers do not appear to consistently take into account that speakers' communicative goals' may shift as a function of context.

Keywords: factive presuppositions, common ground, degrees of belief, communicative strategies, assertion, epistemic contextualism, knowledge norm of assertion

1 Introduction

When communicating propositional content, interlocutors negotiate knowledge and work towards expanding their shared common ground, the sum of the information that is assumed to be true by the interlocutors of a conversation (Stalnaker, 1974). When speakers convey content, they merely propose propositional content to be added to the common ground and whether that content is actually accepted is decided in the subsequent discourse. The way content is negotiated may be affected by the way speakers propose that content. Considering (1), a speaker uttering *believe* (1a) might only have plausible reason for thinking that the door was locked, whereas *know* (1b) or the bare assertion (1c) seem to convey some degree of certainty about the matter.

- (1) a. I believe that the door is locked.
- b. I know that the door is locked.
- c. The door is locked.

One may assume that a speaker who is highly certain of a proposition being true and wants to maximise the chances of it being added to the common ground would choose a formulation expressing high levels of certainty such as *know* or the bare assertion. However, it has been widely disputed what kind of epistemic state – knowledge or (weak) belief – induces speakers to assert content via the bare assertion and, in turn, what epistemic state is conveyed by such a formulation to hearers. Likewise, it is yet to be determined under which circumstances speakers would choose to produce maximal certainty, since expressing maximal certainty may not always aid speakers’ communicative goals.

This paper examines speakers’ production choices and hearers’ interpretations of the formulations *believe/know*/bare assertion. We test (i) how (maximal) certainty is expressed and inferred; and (ii) whether speakers adjust their production choices depending on their communicative goals; and if so whether hearers are able to adjust their interpretations correspondingly. For this purpose, we created two contrasting scenarios – one with a cooperative interlocutor and one with a potentially uncooperative interlocutor. These scenarios were used in two experiments that measured production and comprehension respectively. Below we review arguments for the relative epistemic certainty associated with *believe*, *know*, and the bare assertion. We then present two experiments that test (i) and (ii) for the communication of different levels of certainty across cooperative versus uncooperative contexts. The results suggest that *know* is epistemically the most powerful formulation – stronger than the bare assertion which implies that the norm of assertion is one of weak belief – but that the bare assertion may still be preferred over *know* for expressing maximal certainty in cooperative scenarios.

2 Epistemic states and the common ground

2.1 The Knowledge or Belief Norm of Assertion

Assuming a Stalnakerian view about the communication of content, speakers are taken to propose a proposition to be added to the common ground, where the common ground is the set of propositions that are presupposed – i.e. mutually taken for granted – in the conversation (Stalnaker, 1974). Content may be proposed in different ways, with one locus of variation being at the epistemic level: speakers may express varying degrees of belief about a proposition being true and thereby signal the extent to which propositional content should be negotiated further. Expressing their degrees of belief may be crucial in cooperative communication during which interlocutors jointly work towards expanding their common ground. Here, interlocutors wish to avoid accepting potentially false or controversial propositional content into the common ground. Thus, a speaker who is uncertain about a proposition being true may communicate their uncertainty, e.g. via the usage of *believe*, and thereby allow that the proposed content can be discussed further. In contrast, by expressing high certainty a speaker may signal that further negotiation is not necessary.

Intuitively, high confidence can be expressed by producing *know* or the bare assertion. However, in linguistics and philosophy it has been disputed what epistemic state – knowledge or (weak) belief – speakers require in order to use a bare assertion (and what epistemic state hearers attribute to speakers as a result of their using the bare assertion). Hence, it is an open question whether the bare assertion can be understood to pattern more with *believe* or *know*, both formulations that speakers can produce to explicitly express their epistemic state. Williamson (2000) claims that only knowledge warrants assertion, and hence, one can truthfully assert p only if one **knows** p . This view, termed the Knowledge Norm of Assertion, can capture Moore’s (1942) paradox and the observation that ‘ p but I don’t know that p ’ is incoherent, as well as Unger (1975)’s observation that “How do you know?”, which presupposes that the speaker knows p , is a felicitous response to an assertion of p . The Knowledge Norm of Assertion, following Williamson (2000), can be understood as the prerequisite to truthfully assert p and does not explicitly distinguish the prerequisites of p and *know that p* (to our knowledge). One could conclude that this account extends to *know*, if one assumes that the bare assertion and *know* are logically equivalent: if one is able to assert p , since one knows p one should similarly be able to utter *know that p* . In that case, the bare assertion and *know* would be predicted to pattern together in the way speakers select such formulations based on their degree of certainty about p .

One objection to the Knowledge Norm of Assertion is that it implies that speakers have to go to great lengths to make sure that their proposition is correct before asserting it. Taking knowledge

to be a prerequisite for assertion arguably imposes a condition that is stronger than the one Grice (1975) put forth when formulating the Maxim of Quality. The Maxim of Quality enjoins a speaker not to say what they believe to be false or for which they lack adequate evidence. If adequate evidence would mean being in the position to know that p , utterances about the future, e.g., ‘the sun will come up tomorrow’, would be rendered infelicitous. This argument can be extended to the more general observation that people seem to make ‘weak’ assertions quite regularly without them being rendered infelicitous. As an example consider (2), taken from Mandelkern and Dorst (2022, p. 5):

- (2) a. Mark: What will John bring for dinner?
- b. Liam: He’ll bring Indian. That’s his favourite, after all.

For this example, Mandelkern and Dorst (2022) argue that assuming Liam is 55% confident that John will bring Indian, 25% confident of Chinese, 15% confident of Thai, and 5% confident of Italian, Liam’s utterance seems felicitous despite him being not maximally certain that John will bring Indian.

Correspondingly, some philosophers who do not endorse the Knowledge Norm of Assertion suggest that the prerequisite for asserting a proposition is the same as for strongly believing it. For example, Lackey (2007) argues that asserting p requires the speaker to strongly believe that p , i.e. the speaker requires enough epistemic support to be able to justify their belief. Either way, these arguments suggest that the bare assertion represents a formulation to be reserved for the highest levels of certainty.

2.2 Belief and assertion are weak

Both accounts discussed so far hold that the epistemic state for a speaker to truthfully assert propositional content is either knowledge or some sort of strong belief, and hence the production of the bare assertion should be similar to the production of either *know* or *believe* (in the case of *believe* being used for strong belief). However, neither account explains the availability of three distinct formulations, nor the intuition of naive speakers that there are differences between their communicative effects, specifically in that speakers producing *I believe that p* are taken to be putting forward the proposition p more tentatively than they are when producing p or *I know that p*. Similarly, if p and *know that p* were equivalent in that they both convey that the speaker *knows that p*, this would make the option *know that p* redundant.

More recently, Hawthorne, Rothschild and Spectre (2016) contest the idea that the evidential warrant required for asserting a proposition is the same as the warrant for believing it. Hawthorne et

al. claim that belief is weak – as opposed to the concept of full belief/maximal certainty – which means that thinking that a proposition is likely may entitle a speaker merely to utter *I believe that p* but not to assert *p*. Thus, the authors claim that the standards for the bare assertion are strictly higher than for believing *p*. While Hawthorne et al. capture the intuition that *believe* conveys lower degrees of belief than the bare assertion, the authors maintain the claim that producing the bare assertion requires an epistemic state such as knowledge or maximal certainty. Such a claim runs into similar problems to those outlined above for Williamson (2000)’s account, as a knowledge norm seems too strong for the bare assertion.

In contrast, Mandelkern and Dorst (2022) argue that not only *believe* but also the bare assertion requires weak belief. In this way, they can capture the intuition of interlocutors that speakers very rarely know (in any philosophically defensible sense) that *p* is true when asserting *p*. Thus, whereas asserting *p* does convey at least a weak sense of belief in *p*, assertion does not require the standard of full and complete knowledge. On this account, the authors argue that the standards of assertion vary with context. For example, the production of future-directed statements such as ‘it will rain tomorrow’ is felicitous despite the fact that the speaker may be uncertain about it raining tomorrow, since uncertainty about future-directed statements is already common ground, and thus the epistemic standard for assertion is low in such contexts. Mandelkern and Dorst can similarly capture past-directed statements which are preceded by ‘I don’t know that p’ claims, see (3). Since it is common ground that Speaker1 is uncertain about when the war started, it is felicitous to assert (3c).

- (3) a. Speaker1: I don’t know. I know it was in the 1700s, but I don’t know more than that.
 b. Speaker2: Okay, just have a guess.
 c. Speaker1: Ok, hm. The war started in 1760.

Mandelkern and Dorst’s account suggests that whereas the ordinary usage of *believe p* conveys a weak attitude towards *p* and may be produced by speakers with relatively low degrees of belief about *p* being true, the bare assertion may capture broader levels of attitudes such that a speaker asserting *p* may do so with relatively low as well as high degrees of belief. In contrast, *know* expresses knowledge of *p* or maximal certainty. They furthermore make the case that context plays a role such that if uncertainty is already common ground, speakers are not obliged to make their uncertainty explicit by using formulations such as *believe*.

2.3 The hedging device *believe*: *believe* < bare assertion?

A further argument for *believe* being associated with weak belief/lower degrees of belief than the bare assertion comes from politeness theory. This perspective is crucial, especially in cooperative

communication where interlocutors are expected to adhere to politeness constraints. More specifically, the discourse effects of *believe*, along with other uncertainty expressions such as *probably* and *might*, have been contrasted in the literature with the bare assertion. Whereas a speaker who asserts p proposes to add p to the common ground, uncertainty expressions can be used as *down-toners* (Holmes, 1982) or *hedges* (Lakoff, 1973; Fraser, 1975; Brown & Levinson, 1978) to weaken the assertive strength of p , yielding discourse effects such as politeness.¹ For example, instead of straight out asserting a proposition as in (4a), a speaker might want to gently introduce the subject that the cook did not season their soup well enough and thereby hedge the assertive strength (4b) to come across as more polite. Crucially, for *believe* to be used as a hedging device, one has to assume that *believe* is associated with weaker epistemic states than the bare assertion.

- (4) a. Your soup needs more salt.
 b. I believe your soup needs more salt./Your soup might need more salt.

Politeness is often analysed through the lens of facework, i.e., through interlocutors' aims of managing their positive and negative face (Brown & Levinson, 1987). Whereas positive face reflects the interlocutors' maintenance of a positive self-image, negative face reflects the interlocutors' freedom to act on their own terms. Face-threatening actions can damage the face of either the speaker or hearer. For example, in a scenario where two colleagues are discussing a deadline, as in (5), Colleague2 may be threatening Colleague1's positive face by disagreeing with their statement. Similarly, this situation could be potentially face threatening for Colleague2 because they could come across as uncompassionate. A strategy to lessen the severity of the threat would be the use of an uncertainty expression such as 'believe' (5b) – i.e., hedging – instead of asserting the criticism directly (5a).

- (5) Colleague1: I was so relieved when I realised that the deadline is tomorrow.
 a. Colleague2: The deadline was yesterday.
 b. Colleague2: I believe that the deadline was yesterday.

Empirical research suggests that speakers indeed consider motives such as politeness when communicating their degrees of belief. For example, Juanchich and Sirota (2015) investigated the communication of more or less face-threatening news to a friend. When participants were tasked with conveying news about unfortunate events (e.g., a friend's bad investment) which were characterised

¹As was raised by an anonymous reviewer, hedging may not necessarily be employed by speakers to be polite but instead to be defensive and to avoid being held responsible for possible false information. For example, work by Ochs (1976) suggests that speakers hedge their statements to avoid accountability and unpleasant social consequences that may result from their communicative contribution. In that sense, speakers would hedge their statements out of relatively selfish rather than cooperative reasons. Thus, future research is necessary to tease apart hedging as cooperative versus self-centred communicative tool.

as being 50% likely to occur participants hedged their utterances, understating their confidence by using expressions such as ‘a very small probability’, ‘a small probability’, or ‘slightly probable’.

From the hearers’ perspective, [Bonneton and Villejoubert \(2006\)](#)’s findings have shown that participants tend to judge severe negative outcomes to be more likely than the speaker conveys. Taken together, these studies suggest that speakers downplay the probability of severe adverse events but that hearers are aware of this and compensate accordingly in their interpretation. In general, speakers who use hedging strategies are perceived as less authoritative and confident than those who do not ([Hosman, 1989](#); [Crismore & Kopple, 1997](#)) but at the same time, such speakers are also perceived as warmer.

This section suggests that, applied to the three formulations under investigation, speakers tend to do more than consider their own epistemic state when choosing between the production of the bare assertion versus *believe*. Accounts which consider that speakers may hedge their statements would predict that while speakers use *believe* for overall lower degrees of belief than the bare assertion, they may similarly use *believe* for relatively low degrees of belief but sometimes even for relatively high degrees of belief due to politeness considerations. Hedging may be helpful in cooperative situations in which interlocutors have different opinions but want to reach a shared common ground: speakers who engage in polite behaviour via hedging may signal their willingness to discuss their conveyed content further as well as their consideration for the beliefs of the other interlocutors.

2.4 The presupposition trigger *know*: bare assertion < *know*?

A further argument for *know* being associated with higher degrees of belief than the bare assertion comes from work on common ground and discourse updating, in which *know* is widely argued to trigger a factive presupposition. In (6–7), the speaker presupposes that the deadline is tomorrow. This content can be identified as a presupposition of (6) and (7) on the basis of its ability to project from under the scope of operators such as negation: despite (7) being negated, in both examples, the same content (that the deadline is tomorrow) is presupposed. Thus, the presupposition is argued to project from under the scope of negation.

- (6) She knows that the deadline is tomorrow.
- (7) She doesn’t know that the deadline is tomorrow.

Characteristically, presuppositions are already shared knowledge ([van der Sandt, 1992](#)) and in the common ground of the involved interlocutors ([Stalnaker, 1974](#)). However, presuppositions can be used to introduce new information. Following [Lewis \(1979\)](#) and [von Stechow \(2008\)](#), hearers who lack

the shared knowledge may accommodate the presupposition: speakers act as though the presupposition is already part of the common ground, and hearers respond by adjusting their world-view to incorporate that presupposition. Crucially, accommodation is based on the assumption that hearers prefer to accept the presupposition rather than challenge it. This generally holds for cooperative communication in which presupposed content is sincerely taken by the speaker to be uncontroversial.

The factivity of propositional content when introduced via *know that p* may contribute to *know* being associated with a higher level of certainty than the bare assertion and make *know* a more powerful discourse move. Accounts assuming that *know* is associated with higher epistemic strength would predict that speakers use *know* for higher degrees of belief than the bare assertion and hearers infer *know* to convey higher degrees of belief than the bare assertion. However, as was briefly mentioned, presupposing new content that is potentially controversial and treating it as already being part of the common ground could be seen as an uncooperative discourse move (Schwarz, 2019). For this reason, it's possible that speakers only use *know* as a preferred alternative to the bare assertion when communicating controversial content that they'd prefer to remain uncontested in an uncooperative communicative context. For the purpose of this paper, we will focus on the claim that presupposing controversial content may be considered uncooperative and leave it to future research to investigate whether speakers can use *know* strategically to sneak content into a conversation.

Much work in pragmatic theory assumes that speakers are usually cooperative, and this assumption helps explain the emergence of additional contextually-determined meanings. The question naturally arises of how pragmatic reasoning is supposed to work in an 'uncooperative' context.² At the level of assertion, it is clear that interlocutors can understand people perfectly well without making any assumptions about their accuracy or trustworthiness as interlocutors – that is to say, interlocutors can recognise or fail to recognise an assertion as a falsehood while still being able to understand the propositional content of the assertion. Crucially, even at the level of pragmatic phenomena, it is possible to deceive someone by implicature, as discussed by Meibauer (2014). On the empirical side, Franke, Dulcinati and Pouscoulous (2020) found that speakers expect their hearers to infer implicatures even in contexts where it is known that the speakers are perceived as uncooperative. Uncooperative behaviour in the form of violating the maxims of Relevance and Informativeness has however been found to come with social costs (Beltrama & Papafragou, 2023). In effect, speakers can be superficially cooperative, even to the extent of their utterances giving rise to implicatures, while being uncooperative at a deeper level, in the sense that they are deceptive, attempting to give their hearers an inaccurate picture of reality. We take it that the scenario we study involves uncooperativity only at this deeper level.

²We thank an anonymous reviewer for raising this point.

3 The present study

This study aims to investigate speakers' and hearers' epistemic considerations for the use of *believe p/know p/p* (the bare assertion) when two different communicative goals are at play. Experiment 1 is a production experiment where participants chose between *believe p/know p/p* in two contrasting scenarios, i.e. a cooperative versus an uncooperative scenario. Participants also estimated their confidence in *p* based on available evidence. Experiment 2 is a comprehension experiment investigating the epistemic states that hearers attribute to speakers when interpreting *believe p/know p/p*. We test whether hearers infer speakers' epistemic states for the use of *believe p/know p/p* in accordance with our production results, and whether hearers are sensitive to any scenario-driven adjustments by speakers.

To investigate the epistemic states associated with *believe*, *know* and the bare assertion we will consider two scenarios with different communicative goals. In one, which we consider cooperative, the speaker wishes to discuss their knowledge state with their interlocutor, adhering to politeness constraints, to reach a jointly-shared communicative goal. In the other, which we consider uncooperative, the speaker does not intend to give their interlocutor full information about their knowledge state and does not need to adhere to politeness constraints. In this way we will be able to establish whether a relative epistemic ranking of the three formulations holds across different communicative scenarios. In order to create a coherent cover story where it is plausible for a speaker to engage in cooperative and uncooperative communication we used a detective cover story from previous experiments (Lorson, Cummins & Rohde, 2021) with two different scenarios: a briefing with a colleague (cooperative scenario) and an interrogation of a suspect (uncooperative scenario).³

We understand a briefing with a colleague to be a semi-standard scenario where interlocutors jointly work towards a communicative goal.⁴ We understand an interrogation to be uncooperative, since an interrogator cannot assume that the suspect (who is potentially guilty) works towards the same communicative goal, i.e. finding the truth about what happened in the criminal case. The suspect most likely does not want to disclose all information they have in order to avoid being compromised in any way. Thus, both interrogator and suspect are engaging in a reasoning process about what the other knows and both may be using interrogation strategies and counter-interrogation strategies such as bluffing by overstating their certainty in order to reach/hide the truth (Granhag & Hartwig,

³An anonymous reviewer pointed out that our 'uncooperative' scenario may well be described as an 'adversarial' scenario, since interlocutors may fail to be cooperative even in contexts with shared communicative goals. 'Adversarial' therefore emphasises the nature of the contexts in which speakers deceive or lie.

⁴We describe the briefing as a semi-standard scenario to acknowledge that engaging with pieces of evidence may be a deviation from a standard briefing scenario. We thank an anonymous reviewer for clarification.

2008; Luke, 2021).

In the production experiment, based on Lorson et al. (2021), participants are shown evidence that bears upon the likelihood of the proposition under discussion and then asked to choose between utterance options with which to express their epistemic attitude towards that proposition. We used pieces of evidence as a way of enabling participants to arrive at subjective estimates of event probabilities. We assessed participants' degrees of belief after the production task by asking them to rate how likely they thought each event was to have taken place given a piece of evidence. In the comprehension experiment, participants are asked to estimate how certain a speaker is given what they have uttered, in the context of the earlier production task. The full set of data for both experiments and the experimental materials and R scripts are available here: [link for review](#).

3.1 Experiment 1: Production

The first experiment investigated the production of *know*, *believe* and the bare assertion in two controlled, contrasting scenarios. The goal of this experiment was to examine speakers' production choices of the formulations *believe/know*/bare assertion to test how maximal certainty versus weak belief is expressed and to test whether speakers adjust the way they introduce content depending on the communicative setting.

The experiment consisted of two tasks. For the main task, participants were asked to play the role of a detective in an investigation of an art heist where they brief a colleague and interrogate a suspect, relying on pieces of evidence about the suspect's whereabouts. In this way we elicited expressions that correspond to degrees of belief. After the production task, participants then evaluated their confidence in each piece of evidence retrospectively. We refer to this evaluation as the evidentiality measure. Thus, participants choose formulations concerning propositions about which they have a range of confidence levels (low to high) and they do so in two scenarios that differ in their assumed cooperativity (briefing vs interrogation).

3.1.1 Participants

Participants (N=86) were recruited from the crowd-sourcing platform Prolific. We recruited only those with an approval rate above 90. Participants were paid an average of £7.60/hour (the average duration of the experiment was 25 minutes). The age of the participants ranged from 19 to 66 years, with a mean of 35 years (median = 33, mode = 26). 63 participants stated their preferred pronoun as she/her, 22 chose he/him, 0 chose they/them, and 1 person chose not to disclose this.

3.1.2 Design and Materials

For the production task, each participant was exposed to both scenarios, briefing and interrogation. The order of scenario was counterbalanced across participants. Each critical item was presented to each participant once, either in the briefing or interrogation scenario, and paired either with evidence that we considered to be weak in evidentiality or with evidence that we considered to be strong in evidentiality. This yielded a 2×2 design (scenario \times strength). However, for the analysis, we did not rely on this categorisation of evidential strength, but instead on the participants' evaluation of the evidence in the evidentiality rating post-test (see below). To expose participants to both scenarios, we introduced two suspects. In this way, participants interrogated suspect1 and briefed a colleague about suspect2 or the other way round. This yielded 40 unique items: 10 briefing items per suspect and 10 interrogation items per suspect, each accompanied by either weak or strong evidence. Participants consequently saw 20 of these items – 10 briefing items and 10 interrogation items, each scenario being about a different suspect. In each scenario block, 20 filler items were used, of which 10 were control items which functioned as attention checks. Within each scenario block, the order of the items was randomised.

The critical items consisted of a picture containing a question/answer pair and a picture of a piece of evidence. The picture with the question/answer pair set the scene for the scenario manipulation: either participants saw a picture of a briefing room facing a colleague, or they saw an interrogation room where they would be confronted with a suspect; see Figures (1 and 2) and [Appendix III](#) for full details.⁵

⁵To be clear, all participants saw the same colleague, but either suspect1 or suspect2. We did not find any systematic differences in how participants reacted towards the two suspects.

Did Emily Brown have any financial problems?



Evidence:

HARRIOT BANK

Statement Ending 30/11/2019
Page 1 of 4

RETURN SERVICE REQUESTED

Emily Brown
2/1 South Bridge
EH1 1LL Edinburgh

Managing Your Accounts

- Primary Branch: Edinburgh
- Phone Number: XXXXX XXXXXX
- Online Banking: HarriotBank.com
- Mailing Address: 2/1 South Bridge EH1 1LL, Edinburgh

Summary of Accounts

Account Type	Account Number	Ending Balance
HARRIOT RELATIONSHIP CHECKING	XXXXXXXX4101	£ -15,045

Primary Checking

Account Summary		Amount
Date	Description	
01/11/2019	Beginning Balance	£ -9,000
30/11/2019	Ending Balance	£ -15,045

Figure 1: Sample experimental item for the production experiment: Briefing, strong evidence, with colleague interlocutor, discussing suspect Emily Brown.

Did you have any financial problems?



Evidence:

Statement of Emily Brown's friend 20/12/2019
Name: Greta Lawson
Born: 01.02.1982
Issue: Emily Brown's financial situation

Campbell: "Did Emily Brown mention anything about her financial situation?"
Greta Lawson: "She mentioned that she hopes to win in the lottery to buy that house in the south of France that she wanted for so long."

Figure 2: Sample experimental item for the production experiment: Interrogation, weak evidence, with interlocutor suspect Emily Brown.

The question in the heading of the picture was meant to be one already posed by the participant to the addressee (either the colleague, Figure 1, or the suspect, 2). The answer to the question was provided by either the colleague or suspect, depending on the scenario, in the form of a speech bubble. The participants were asked to react to the colleague's/suspect's answer by selecting one of the formulations *I know that p*, *I believe that p* and the *p* (the bare assertion), see (8) for a briefing and (9) for an interrogation item. Alternatively, they were able to choose the option *other* and formulate their own utterance. The order in which the expressions were displayed was randomised for each trial.

- (8) Briefing item (corresponding to Figure 1):
Detective (Participant): Did Emily Brown have any financial problems?
Colleague: Financially the suspect was doing alright.

Detective (Participant):

- a. I know that the suspect was in need of money.
- b. I believe that the suspect was in need of money.
- c. The suspect was in need of money.

(9) Interrogation item (corresponding to Figure 2):

Detective (Participant): Did you have any financial problems?

Suspect: Financially I was doing alright.

Detective (Participant):

- a. I know that you were in need of money.
- b. I believe that you were in need of money.
- c. You were in need of money.

The response *other* was excluded from the analysis. We included *other* mainly to give participants more freedom in their production choice and to create a more natural experience. The response *other* made up 13.5% of the data.⁶

The manipulated pieces of evidence ranged from pictures to witness statements. As was mentioned above, the evidence was roughly manipulated to be weak or strong but the analysis will rely on the participants' evaluation of the evidence in the analysis. For example, for the items (8) and (9) a bank statement was provided as strong evidence (Figure 1) and a statement of a friend who mentions potential financial difficulties as weak evidence (Figure 2); see Appendix III for full details.

The filler items were turns between the suspect/colleague and the participant that had nothing to do with the case. For the 10 control items, the information was provided in the picture (Figure 3) and the participant had to choose the correct answer (here 11am), see (10). The option *other* was also available.

⁶We did not norm the exchanges for acceptability/naturalness. However, in cases where participants did not like the response options they could always type in their own response by choosing the option *other*. That 13.5% of the data were *other* responses suggests that participants were generally satisfied with the production options. When participants chose *other*, in the briefing, they mainly chose to produce utterances containing alternative uncertainty expressions other than *believe* such as *Johanna Smith may have entered via the air ventilation ducts.* or referred to the evidence, e.g. *Johanna was scheduled to meet Emily Brown on that day.*. In the interrogation, participants tended to refer back to evidence, e.g. *We have an email showing she contacted an art dealer.*, *The locksmith says it was a high security institutional key. . .*, and/or asked follow-up questions, e.g. *But you bought a pass for the whole day, why was that?*.

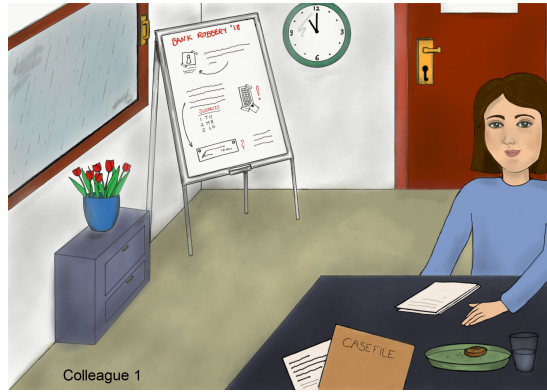


Figure 3: Control item/ attention check Briefing in the production experiment

- (10) Detective (Participant): Oh look at the time. Is it already _____?
 [11am|noon|2pm]

After the main task, participants proceeded to the evidentiality rating post-test where they were asked to evaluate the pieces of evidence they had seen in both scenarios: *Given the piece of evidence below, how certain are you that p?*, where p is the complement proposition from earlier in the experiment. In the above example, this would have been either the bank statement (strong evidence) or the statement of a friend (weak evidence), depending on which piece of evidence they had seen in the preceding scenario. Participants saw 20 pieces of evidence in total and each piece of evidence dealt with a separate proposition. To communicate their certainty, participants adjusted a slider from 0 (not at all certain) to 100 (very certain).

3.1.3 Procedure

First, participants were asked to give informed consent to take part in a fictional investigation of an art heist in the role of a detective. We then informed them about the structure of the experiment which was as follows: (1) production task, engaging in two discussions, (2) evidence rating, (3) demographic questionnaire. After that we introduced participants to the task in the form of a story about an art heist in Edinburgh involving two suspects. Since the lead detective on the case had gone missing, the participants were asked to help out with solving the case. We introduced both scenarios by stating that one of the suspects had been arrested and participants were then asked either to prepare for the interrogation of the suspect with a colleague (briefing scenario) or to interrogate the suspect right away (interrogation scenario). After the trials for the first scenario, participants were then notified about a ‘breakthrough’ in the case, i.e. that a second suspect had been captured which required another briefing or interrogation depending on which scenario the participant had previously participated in. We instructed participants to converse with the colleague/suspect about

different topics, including questions about the case, by choosing between production choices offered to them. They were also told that for parts of the interaction they would need to look at the evidence that had been collected. In the briefing scenario participants were asked to find the best way to help their colleague, and in the interrogation scenario to find the best way to interrogate the suspect. After having completed the production task participants were asked to rate the quality of the pieces of evidence they had seen. More specifically, participants were asked how certain they were that a specific event happened (e.g. the suspect being at a party) given a piece of evidence (e.g. stills of video footage of the suspect being at the party). The evaluation task was followed by the voluntary demographic questionnaire.

3.1.4 Predictions

Depending on whether the norm of assertion is knowledge (KNOWLEDGE NORM PREDICTION) or belief (BELIEF NORM PREDICTION), the bare assertion should pattern with *know* or *believe*. More specifically, averaged over scenarios, participants are predicted to either choose the bare assertion similarly to *know* to express maximal certainty (i.e. high evidentiality ratings) or to choose the bare assertion similarly to *believe* to express lower certainty (i.e., lower evidentiality ratings).

Another possibility is that weak belief is the norm of assertion but the standards of assertion are context-dependent (CONTEXT-SENSITIVE PREDICTION). In such a case, averaged over scenarios, participants are predicted to choose the bare assertion for a wide range of evidentiality ratings, depending on the context (the experimental item). This is because a participant may take a specific context to be one in which uncertainty is already part of the common ground, and hence take the standard of assertion to be weak belief, or one in which it is not, in which case they are predicted to take the standard of assertion to be something like knowledge.⁷ Overall, if the CONTEXT-SENSITIVE PREDICTION applies, the bare assertion should be on average used for higher evidentiality ratings than *believe*, and lower evidentiality ratings than *know*. In contrast, *believe* and *know* should be used for a smaller range of relatively low and relatively high evidentiality measures, respectively.

If *believe* is epistemically weaker than the bare assertion, *believe* may be utilized by speakers as a hedging device. Assuming that in a cooperative scenario interlocutors are working towards a shared communicative goal which involves expanding their common ground, conforming to politeness norms via hedging may aid the negotiation of content: speakers may hedge their statements to signal their willingness to discuss their conveyed content further as well as their consideration

⁷Note that we did not design our experiment to directly test these context effects. We are only able to predict a greater variability in the production of the bare assertion versus *believe* and *know* assuming that participants take item-specific contextual uncertainty into consideration across our two scenarios.

for the beliefs of the other interlocutors. Thus, in the cooperative scenario as opposed to the uncooperative scenario, participants are predicted to use *believe* more frequently, and to do so for higher evidentiality ratings due to politeness considerations.

Similarly, if *know* is epistemically stronger than the bare assertion, the presuppositionality of *know* may play a role in its production. If a speaker who introduces new propositional content via *know* that *p* presupposes the factivity of *p* and thereby treats it as being already shared knowledge, it should be considered uncooperative to use *know* for controversial content. Due to the experimental setup being one in which participants are asked to contradict their interlocutors, the content participants are asked to communicate is controversial per se. We assume, however, that weakly evidenced statements are more controversial than strongly evidenced statements, in terms of their admissibility into the common ground. Furthermore, the wish to avoid uncooperative behaviour should be more relevant in what we have termed the cooperative scenario. Thus, participants are predicted to use *know* less frequently in the briefing than in the interrogation and use *know* for higher evidentiality ratings in the briefing than in the interrogation.

3.1.5 Analysis

The 3-way outcome (*know/believe/bare*) was analysed by fitting a Bayesian categorical regression model with varying by-item and by-participants intercepts and slopes using the R (R Core Team, 2020) package *brms* (Bürkner, 2018) which provides an interface to fit Bayesian mixed models using Stan (Stan Development Team, 2017). The experimental factor SCENARIO (briefing/interrogation), the continuous variable EVIDENTIALITY and their interaction⁸ were included to predict the probability of choosing *believe/bare* assertion over *know*, which makes *know* the reference category of the model. The factor SCENARIO was sum-coded: -1 for interrogation and 1 for briefing. EVIDENTIALITY was standardised, such that the variable was centred at zero with a standard deviation of 1. The model included varying intercepts and slopes for participants and items, assuming that the effect of SCENARIO and EVIDENTIALITY on the participants' production choices varies between participants and items.

We used weakly regularising priors, which allowed a reasonably wide range of parameter values and penalised very extreme values. The priors for the intercepts were normal distributions with mean 0 and standard deviation 10. This means that we could be 68% certain that the by-expression

⁸The interaction was not part of the pre-registered analysis. In order to account for possible differences in the way context affects the speakers' production differently depending on the speakers' degrees of belief we will include the interaction in our model. For conciseness we will only discuss the output of the interaction model in the main text. We have included the output for the main effects only model in the Appendix 7.2.

intercepts would fall within -10 and 10 on the log-odds scale. For both fixed effects, normal priors with a mean of 0 and a standard deviation of 1 were used. Random effects were modelled as a correlation matrix and a vector of standard deviations. The standard deviations were assigned half-normal priors with a mean of 0, and a standard deviation of 1. For the correlation matrix, a LKJ(2) prior was used such that smaller correlations are favoured over extreme values such as +/- 1 (Stan Development Team, 2017; Sorensen, Hohenstein & Vasishth, 2016). We ran a second model with uninformative priors that yielded similar posterior distributions to the model with weakly regularising priors; see 6.1 for the output.⁹

Unlike the frequentist analysis, the Bayesian analysis will not produce point estimates but instead posterior distributions over parameters quantifying the probability of each possible parameter value given the data. We will report the posterior mean $\hat{\beta}$ and the 95% credible interval (95%-CrI). 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. We interpret the evidence as reliable if zero lies outside the parameters' 95% credible interval (Kruschke, Aguinis & Joo, 2012).¹⁰

3.1.6 Results

All participants were above 97% accurate for the control items, which suggests that they paid attention during the experiment. Overall, the *believe* was used most often followed by the bare assertion and *know*, see Table 1 for details.

Response	counts
know	328
bare assertion	424
believe	735

Table 1: Raw counts for each formulation of the production experiment ordered from least frequent to most frequent.

In the evaluation task, participants assessed the evidence by using the whole range of the slider: the evidentiality ratings ranged from 0 to 100, with a mean of 74.5. The means, standard deviations

⁹ An anonymous reviewer pointed out that, since suspects reappeared over time, participants might have built up beliefs about each suspect and the likelihood they actually committed a crime, and thus assigned higher evidentiality measures as the experiment progressed. We conducted an additional statistical analysis where we extended the main effects model by including the order in which the pieces of evidence appeared in the second part of the study as a fixed effect. We found no evidence that participants assigned higher evidentiality ratings as the experiment progressed; see Appendix 7.1.

¹⁰For all models in this paper, samples were drawn from the posterior distributions of the model parameters using the NUTS sampler (Hoffman & Gelman, 2013). We ran four sampling chains, each collecting 4000 iterations and discarding the first 1000 iterations as part of the warm-up phase leading to 12000 iterations available for analysis. Chains mixed well (all $\hat{R} = 1.0$).

and medians of the evidentiality measures for the scenarios briefing and interrogation were almost identical (mean=75/74, sd=24.6/25.7, median=80/80.5 respectively) which accorded with the aims of the experimental design. The means, standard deviations and medians of the evidentiality measures for the evidence presented in the briefing and interrogation scenarios were almost identical. A difference in evidentiality ratings between scenarios would have introduced a confounding variable.

With regards to differences between formulations, participants used *know* and the bare assertion in conjunction with evidence that they subsequently judged as convincing (mean evidentiality ratings were 86 and 79.5 respectively), whereas they used *believe* for weaker evidence (mean evidentiality rating was 66.5); see Table 2 for details.

Response	Scenario	mean	sd	median	min	max
know	Overall	86.2	18.8	94.5	0	100
know	Briefing	88.9	15.6	100	19	100
know	Interrogation	84.6	20.3	91.0	0	100
bare assertion	Overall	79.5	23.1	85.5	0	100
bare assertion	Briefing	82.4	21.5	90.0	0	100
bare assertion	Interrogation	76.9	24.2	83.0	0	100
believe	Overall	66.5	26.0	70.0	0	100
believe	Briefing	67.5	25.4	71.0	0	100
believe	Interrogation	65.0	26.8	69.5	0	100

Table 2: Mean evidentiality ratings by formulation and scenario for the production experiment.

The categorical regression model showed that for both *believe* ($\hat{\beta} = -1.58$, CrI:[-1.96, -1.24]) and the bare assertion ($\hat{\beta} = -0.37$, CrI:[-0.69, -0.04]) an increase in evidentiality corresponded to a decrease in log-odds of production; see Table 10. This suggests that, averaging over scenarios, speakers become increasingly likely to choose *know* (over *believe* or the bare assertion) as their certainty increases about the content they want to communicate.

Focusing on the scenario manipulation, for average evidentiality *believe* ($\hat{\beta} = 0.65$, CrI:[0.42, 0.90]) and the bare assertion ($\hat{\beta} = 0.28$, CrI:[0.04, 0.52]) were less likely to be chosen over *know* in the interrogation and more likely to be chosen over *know* in the briefing scenario, and this effect is reliable. The credible intervals of the interaction coefficients for both *believe* ($\hat{\beta} = -0.21$, CrI:[-0.50, 0.07]) and the bare assertion ($\hat{\beta} = 0.02$, CrI:[-0.25, 0.31]) include 0 which suggests that the interaction effect is not reliable.

To place these estimates in a context, our results suggests that at a very high level of certainty (evidentiality = 99.5, which is 1 standard deviation above the mean), speakers are predicted to

Formulation	Coefficient	posterior mean	Est.Error	l-95% CrI	u-95% CrI
believe	Intercept	1.25	0.22	0.82	1.68
bare assertion	Intercept	0.26	0.17	-0.09	0.58
believe	Evidentiality	-1.58	0.18	-1.96	-1.24
bare assertion	Evidentiality	-0.37	0.17	-0.69	-0.04
believe	Scenario	0.65	0.12	0.42	0.90
bare assertion	Scenario	0.28	0.12	0.04	0.52
believe	Evidentiality * Scenario	-0.21	0.15	-0.50	0.07
bare assertion	Evidentiality * Scenario	0.02	0.14	-0.25	0.31

Table 3: Population-level estimates of the categorical regression model with interaction in log-odds with the standard errors and 95% credible intervals. In the table the by-expression intercepts are listed first, then the estimates for the evidentiality effect followed by the estimates for the scenario effect and interaction coefficients. Slope coefficients whose 95% credible intervals do not include zero and are therefore treated as reliable effects are highlighted in bold. The effect scenario is the change in log-odds for the briefing (-1 interrogation, 1 briefing). \hat{R} is a convergence diagnostic which compares the between- and within-chain estimates. Values larger than 1 suggest that the chains have not mixed well.

choose *know* over *believe* with a probability of 0.47 in the briefing, and 0.68 in the interrogation. Likewise, speakers are predicted to choose *know* over the bare assertion with a probability of 0.40 in the briefing and 0.64 in the interrogation.¹¹

¹¹Probability estimates are derived from the log odds estimates via the following formula: $\exp(x)/(1+\exp(x))$. The same can be achieved with the built-in function `plogis()` in R.

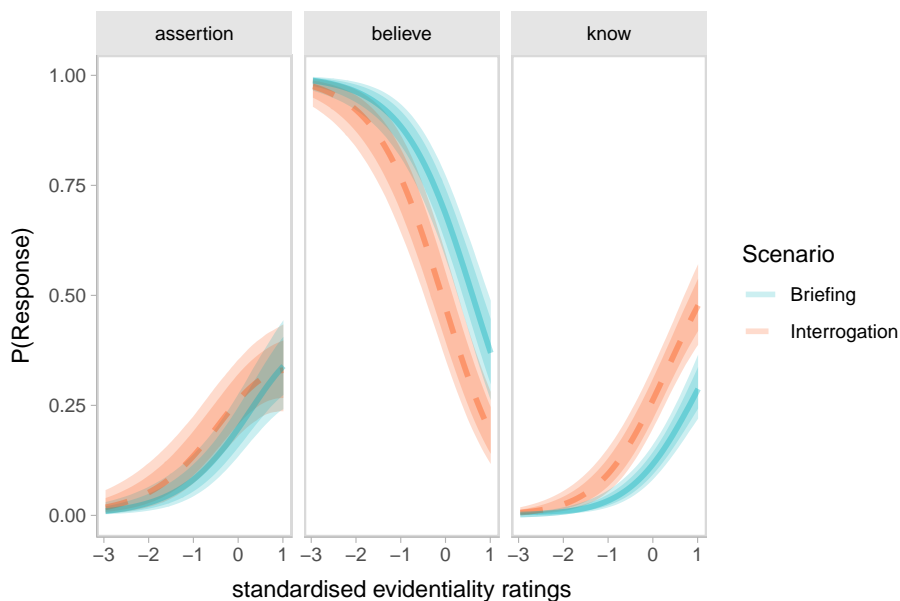


Figure 4: Predictions for the production data. The figure contrasts the two scenarios for each formulation. Log-odds were back-transformed to probabilities (y-axis). The x-axis is the standardised evidentiality measure: 0 stands for an evidentiality of 74.55. An increase of one standard deviation on the standardised scale means an increase of 25 on the original scale. The lines represent the means of the fixed effects, the faded area depicts the 95% credible interval and the darker area the 80% credible interval of the fixed effects.

We visualised the model’s predictions in Figures 4 and 5. In order to facilitate understanding we back-transformed the response data from log-odds to probabilities for both figures. In both figures, the x-axis represents the standardised evidentiality ratings, with zero representing average evidentiality (corresponding to 74.55 on the original scale). The lines represent the means of the fixed effects and the faded area depicts the 95% credible interval and the darker area the 80% credible interval of the fixed effects.

Figure 4 shows the predicted model probabilities of using a specific expression in the contrasting scenarios. With increasing evidentiality, the probability of uttering *know* and the bare assertion is predicted to increase, and the probability of uttering *believe* is predicted to decrease. Focusing on the difference between scenarios, there is little difference in production for the bare assertion. In contrast, the production of *believe* and *know* differs between scenarios to a greater extent: whereas the production of *believe* is more likely in the briefing, the production of *know* is more likely in the interrogation.

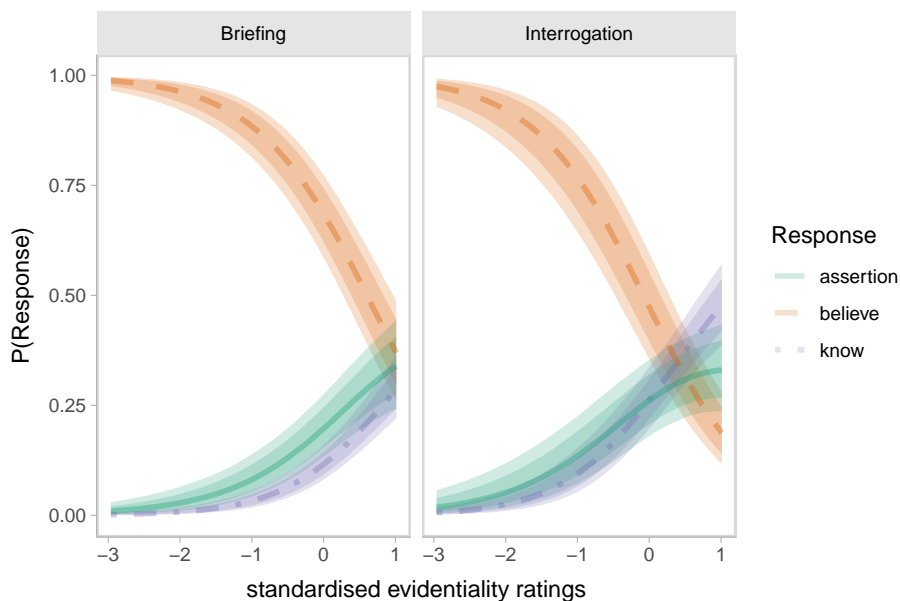


Figure 5: Predictions for the production data. The figure contrasts the three production choices with each other in the two respective scenarios. Log-odds were back-transformed to probabilities (y-axis). The x-axis is the standardised evidentiality measure: 0 stands for an evidentiality of 74.55. An increase of one standard deviation on the standardised scale means an increase of 25 on the original scale. The lines represents the means of the fixed effects, the faded area depicts the 95% credible interval and the darker area the 80% credible interval of the fixed effects.

Figure 5 illustrates the predicted model outcome in a different way than Figure 4 to emphasise the differences between formulations – and specifically between *know* and the bare assertion – in the two scenarios. In the briefing, the model predicts that the bare assertion is more likely to be uttered over *know* for the whole range of evidentiality ratings, and that stronger evidence is needed in order for speakers to choose *know* over the bare assertion. In the interrogation, this tendency remains similar for low evidentiality ratings but the difference between formulations appears less pronounced. Furthermore, Figure 5 suggests that when speakers get more confident – at mean evidentiality ratings 74.55 and above – *know* is the most likely chosen option. Such a trend would mean that the effect of scenario on the speakers’ production of *know* is stronger for high evidentiality ratings than for average and low evidentiality ratings. However, we have to interpret this characterisation with caution, since both interaction terms were not reliable.

In order to understand better how all three formulations are affected by the predictors evidentiality and scenario, we compared possible effect sizes of the predictor scenario on the production of the three formulations for different evidentiality ratings. We did so by calculating differences in

probabilities between uttering a given formulation in the interrogation relative to the briefing for low (1 SD below the mean), average and high evidentiality ratings (1 SD above the mean) from the posterior draws of the interaction model. Thus, for illustration purposes, Figure 6 displays only three reference points of the continuous evidentiality measure. We used the same method as Sós-kuthy and Roettger (2020), and an interested reader may look at our code here [link for review](#).

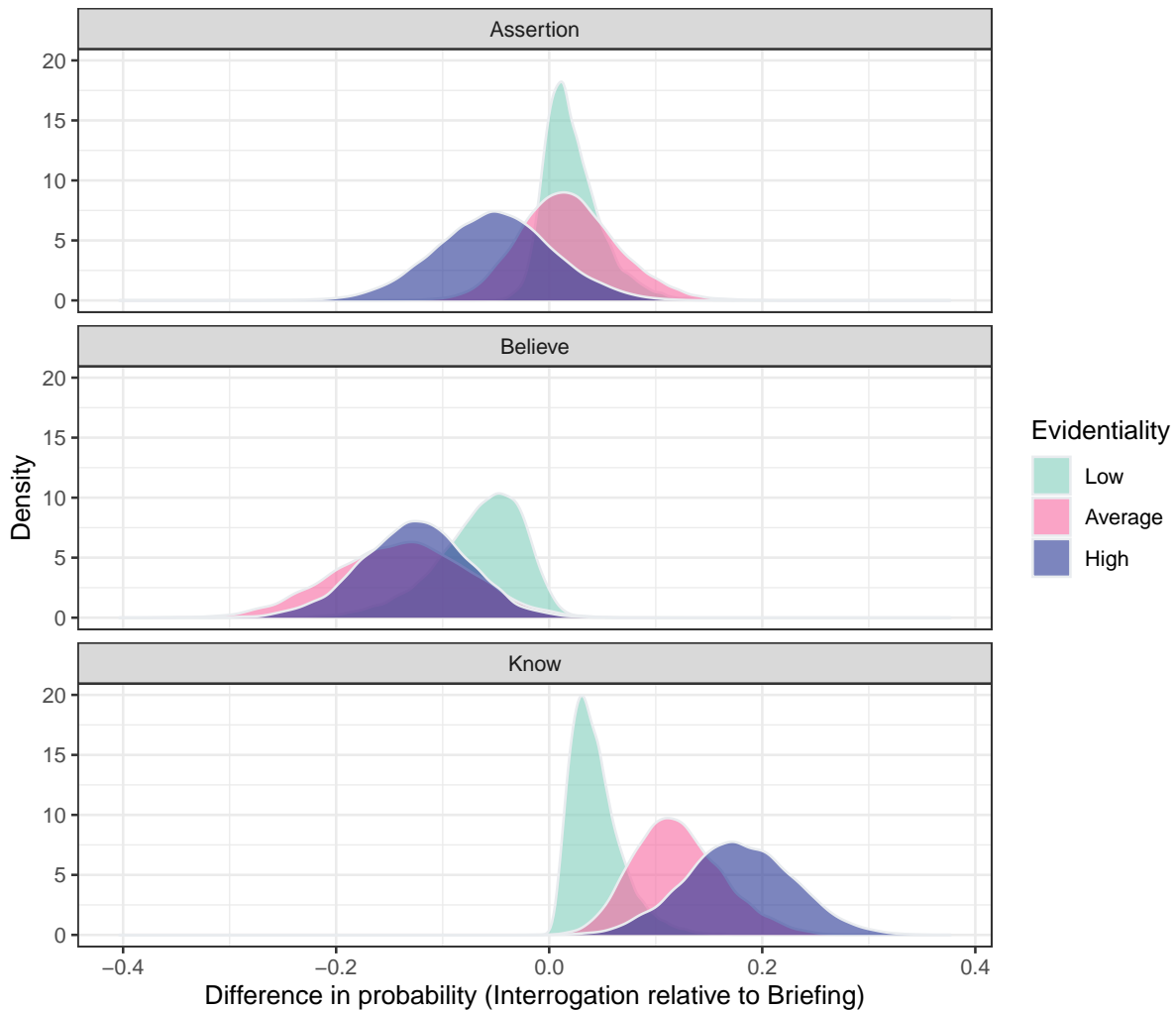


Figure 6: The density plot shows the difference in probability of selecting one of the formulations in the interrogation relative to the briefing for low (green), average (pink) and high evidentiality ratings (blue).

Figure 6 suggests that if there is an effect of scenario on the production of the bare assertion the effect would be tied to high evidentiality ratings meaning that there is a tendency that the bare assertion was more likely produced in the briefing than in the interrogation for high evidentiality

ratings. However, all three density estimates include 0 which means that we only have weak evidence for an effect of scenario on the bare assertion. In contrast, the production of *know* is more clearly influenced by the factor scenario (whereby the effect of scenario is most pronounced for average and high evidentiality ratings), since neither of the density estimates include 0. The figure suggests that, by contrast to the bare assertion, the production of *know* is more likely in the interrogation than in the briefing and gets increasingly likely with increasing evidentiality. The effect of scenario on the production of *believe* is more pronounced for average and high evidentiality ratings, such that the production of *believe* is more likely in the briefing than in the interrogation for average and high evidentiality ratings.

3.1.7 Discussion

We found that participants tended to choose *believe*, bare assertion and *know* for successively increasing evidentiality ratings, suggesting that knowledge or maximum certainty is not required when producing the bare assertion. All in all *know* seems to be the epistemically strongest formulation. The relative ordering of the formulations would agree with the CONTEXT-SENSITIVE PREDICTION. However, averaged over scenarios, all three formulations were chosen for a wide range of evidentiality measures, suggesting that speakers are not always certain about a proposition even when producing *know*.

Regarding the reliable effect of SCENARIO, we found that *believe* was the most frequently produced formulation overall but that participants produced *believe* more frequently in the briefing (absolute count = 425) than in the interrogation (absolute count = 310). Furthermore, *believe* was produced for higher evidentiality ratings in the briefing than in the interrogation. The differences between scenarios for the production of *believe* were numerically strongest for average and high degrees of belief, at which levels *believe* was numerically less likely to be produced in the interrogation than in the briefing.

In contrast to *believe*, the bare assertion was used slightly more frequently in the interrogation (absolute count = 225) than in the briefing (absolute count = 199). The difference in production between scenarios for the bare assertion seems to suggest that participants tended to use the bare assertion more frequently for low and average evidentiality ratings in the interrogation and participants who were very certain about p chose the bare assertion more frequently in the briefing. However, out of all three formulations the production of the bare assertion seems to be the least context-sensitive formulation.

Moving to *know*, in the interrogation, the frequency of producing *know* increased considerably

(absolute counts briefing = 123, interrogation = 205). Context tended to affect high certainties (evidentiality ratings 1 SD above the mean) more strongly: a change from briefing to interrogation meant an increase in probability of *know* being produced particularly for high certainty (above rating 74.55). Thus, participants more frequently chose *know* in the interrogation than in the briefing and with increasing evidentiality this difference tended to get larger.

3.2 Experiment 2: Comprehension

The results of experiment 1 show that speakers produce *believe*, the bare assertion and *know* for successively higher evidentiality ratings. This raises the question of whether hearers infer the same relative ranking. We furthermore wanted to investigate whether hearers take speakers' communicative goals into account while forming an understanding of the speakers' epistemic state. In experiment 2, we test this by eliciting hearers' assessments of speaker certainty in the scenarios used in experiment 1 to explore whether these are based solely on the speakers' production choice (bare assertion vs. *believe* vs. *know*) or also on the communicative setting. The experiment and analysis, specified below, were pre-registered; see [link for review](#).

Participants were asked to play the role of a detective in training and to assist their colleagues in an investigation of an art heist. As in the production task, the investigation involved a briefing and an interrogation. The participants' task was to evaluate their colleague's certainty in these two scenarios when uttering *believe/know* or the bare assertion, by adjusting a slider.

3.2.1 Participants

Participants (N=121) were recruited from Prolific following the approach in experiment 1. Participants were paid £2.86 for a task that was estimated to take 20 minutes (on average the experiment lasted 18 minutes). The age of the participants ranged from 18 to 74 years, with a mean of 38 years (median = 38, mode = 32). 3 participants stated their preferred pronoun was they/their, 76 chose she/her, 40 chose he/him and 2 participants preferred not to disclose a preference.

3.2.2 Design and Materials

The design and materials were used from experiment 1 and were only minimally changed to display an utterance from a second colleague. Each participant was exposed to both scenarios, briefing and interrogation, and the order of scenario was counterbalanced across participants. Each critical item was presented to each participant once, either in the briefing or interrogation scenario, and paired with one of the formulation alternatives yielding a 2×3 design (scenario × formulation). As in the production task, we used pictures to set the scene for the briefing and interrogation: either

participants saw pictures of a briefing room with two colleagues, or they saw an interrogation room with a colleague and a suspect; see Figure (7) for a briefing item where the colleague utters *know* and Appendix IV for full details.



Figure 7: Sample experimental item for the comprehension experiment: Briefing, *know* formulation, with colleagues discussing suspect Emily Brown.

The pictures were assembled like a comic strip. In the briefing (11), Colleague1 answers an already asked question (speech bubble in first picture), and then Colleague2 responds (speech bubble in second picture). In the interrogation (12), the first picture depicts the suspect's answer in the form of a speech bubble and the second picture depicts the colleague's statement. The formulation of the second turn was the critical formulation which was either the bare assertion (11a) or contained *believe* (11b) or *know* (11c).

As in experiment 1, we introduced two suspects in order to ensure that participants saw a particular item only once, in either briefing or interrogation. Thus, one participant would participate in a briefing about Suspect1 and in an interrogation of Suspect2. This yielded 40 unique items:

10 briefing items per suspect and 10 interrogation items per suspect, each accompanied by either weak or strong evidence. Participants consequently saw 20 of these items – 10 briefing items and 10 interrogation items, whereby the briefing items were about different suspects, e.g. if the briefing items were about Suspect1, the interrogation items were about Suspect2. In each scenario block, 20 filler items were used of which 10 were control items which functioned as attention checks. Within each scenario block the order of the items was randomised.

The participants were asked to assess the speaker’s certainty by adjusting a slider ranging from 0 (not at all certain) to 100 (very certain).

(11) Briefing item:

Question: Did Emily Brown have any financial problems?

Colleague1: Financially the suspect was doing alright.

Colleague2:

- a. The suspect was in need of money.
- b. I believe that the suspect was in need of money.
- c. I know that the suspect was in need of money.

(12) Interrogation item:

Question: Did you have any financial problems?

Suspect: Financially I was doing alright.

Colleague:

- a. I know that you were in need of money.
- b. I believe that you were in need of money.
- c. You were in need of money.

The filler items were the same used in the production experiment where participants were asked to engage with the suspect/colleague by choosing a formulation. Again, we included 10 attention checks; see (10) for more details.

3.2.3 Procedure

Before the experiment, participants were asked to give informed consent to take part in a fictional investigation of an art heist in the role of a detective in training. We then informed them about the structure of the experiment, the comprehension task and the demographic questionnaire. We

introduced the task with the same introduction text that was used for the production experiment; see 3.2.3. Both scenarios were introduced by stating that one of the suspects had been arrested. The participants were then asked either to assist two of their colleagues in the briefing, or to assist one colleague in the interrogation of one of the suspects. Participants were informed that they would converse with their colleagues/the suspect about different topics and that their task as a detective in training was to observe and assess their colleagues. The experimental task was followed by the voluntary, demographic questionnaire. The experiment lasted approximately 20 minutes.

3.2.4 Predictions

With this study we are testing the predictions of the previously discussed accounts under the assumption that hearers are rational. More specifically, we investigate whether hearers consider the speakers' degrees of belief and the context (briefing/interrogation) they are in when interpreting utterances. We base our predictions for assessed speaker certainty on the results that were obtained in Experiment 1 whereby hearers are predicted to assign speaker certainties for the three formulations in the following relative order: *believe* < bare assertion < *know*.

Taking the speakers' communicative goals into account, we will formulate the following predictions based on the assumption that hearers are aware of these goals. For the bare assertion formulation, if speakers produce the bare assertion for epistemic states weaker than knowledge/strong belief (WEAK BELIEF NORM OF ASSERTION) and produce the bare assertion at different rates depending on context (CONTEXT-SENSITIVE PREDICTION), hearers are expected to interpret the bare assertion to cover a wider range of speaker certainty values than the other two expressions.

For the use of *believe*, we consider comprehension in cooperative versus uncooperative scenarios and assume that speakers conform to politeness norms via hedging to aid the negotiation of content: in cooperative scenarios, speakers may hedge their statements using *believe* in order to signal their willingness to discuss their conveyed content further as well as their consideration for the beliefs of the other interlocutors. Assuming hearers are aware of the speakers' strategies, participants are predicted to infer higher speaker certainty values for *believe* in the briefing than in the interrogation.

For the use of *know*, we likewise consider its usage in cooperative versus uncooperative scenarios. In cooperative scenarios, speakers should avoid presupposing potentially controversial information and therefore should instead assert rather than risk creating a situation in which the hearer must accommodate potentially controversial content. Such a behaviour may only be relevant in cooperative situations in which interlocutors work together towards a shared communicative goal. Assuming that speakers in turn use *know* more readily – for lower degrees of belief and more frequently –

in the interrogation and hearers are aware of this behaviour, participants are predicted to assign higher certainty values to speakers uttering *know* in the briefing than in the interrogation.

3.2.5 Analysis

The data was analysed by fitting a Bayesian beta regression model with maximal random effects structure. We had preregistered both a linear and a beta model. We decided to proceed with the beta model because (i) the dependent variable was heavily skewed and the linear model’s predictions were less accurate than those of the beta model, and (ii) we didn’t want the model predictions to exceed 0 and 100 (0 and 1 on the probability scale). In order to use a beta regression we transformed the gathered data so that it fits on the open standard unit interval (0, 1) following [Smithson and Verkuilen \(2006\)](#). Here, *S_certainty* is the dependent variable *speaker certainty* and *n* is the number of observations (here 2420).¹²

$$[(S_certainty/100) * (n - 1) + 1/2]/n$$

The experimental factor SCENARIO (briefing/interrogation), FORMULATION (*believe/know*/bare assertion) and their interaction, were included to predict the assessed speaker certainty.¹³ The factor SCENARIO was sum-coded: -1 for interrogation and 1 for briefing. For the factor FORMULATION the following sum coding was used:

	Formulation I	Formulation II
believe	1	0
assertion	0	1
know	-1	-1

Table 4: Sum coding for the predictor *formulation*

The model included varying intercepts and slopes for participants and items, assuming that the effect of SCENARIO and FORMULATION on the participants’ assessment of speaker certainty varies between participants and items.

Again, we used weakly regularising priors to allow for a reasonably wide range of parameter values while penalising very extreme values. The priors for the by-expression intercepts were normal distributions with mean 0 and standard deviation 10. For both fixed effects, we used normal priors

¹²We decided to perform a transformation rather than changing the scale in the experiment to be as close as possible to the way we collected evidentiality measures which was adjusting a slider from values 0 to 100.

¹³This deviates from the pre-registered model which included only main effects. For brevity of this paper we will only discuss the interaction model. See [Appendix 7.4](#) for the main effects model output.

with a mean of 0 and a standard deviation of 1. Random effects were modelled as a correlation matrix and a vector of standard deviations. The standard deviations were assigned half-normal priors with a mean of 0, and a standard deviation of 1. For the correlation matrix, a LKJ(2) prior was used. We ran a second model with uninformative priors that yielded similar posterior distributions as the model with weakly regularising priors, see 6.2 for the output.¹⁴

3.2.6 Results

Overall, participants assessed speaker confidence by using the whole range of the slider. Here, we are looking at the data before being transformed. The speaker confidence ratings ranged from 0 to 100, with a mean of 69.7. The confidence ratings again were similar between the briefing and interrogation, suggesting that the pieces of evidence were not perceived to vary in their evidential strength between scenarios (mean=0.69/0.7, sd=0.25/0.24, median=0.75/0.75 respectively).

Response	Scenario	mean	sd	median	min	max
know	Overall	80.3	21.7	86	0	100
know	Briefing	79.8	22.2	86	0	100
know	Interrogation	80.7	21.2	86	0	100
bare assertion	Overall	72.9	22.3	77	0	100
bare assertion	Briefing	72.0	22.8	77	0	100
bare assertion	Interrogation	73.8	21.7	77	0	100
believe	Overall	56.0	21.8	57	0	100
believe	Briefing	54.2	21.5	55	0	100
believe	Interrogation	57.8	22.1	60	0	100

Table 5: Mean speaker confidence ratings by formulation

Regarding the impact of a speaker’s formulation on the assigned confidence ratings, on average participants assigned the highest confidence to a speaker uttering *know* (80.3), followed by the bare assertion (72.9), and then *believe* (56.0); see Table 5 for details.

¹⁴See footnote 9.

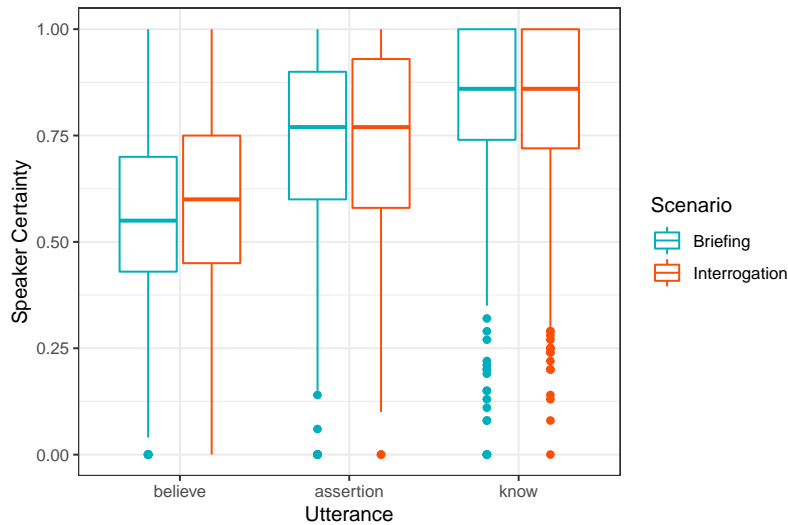


Figure 8: By-expression speaker confidence ratings for each scenario (briefing in blue, interrogation in red). The figure shows the median of assigned quantities (line) and the upper quartile and lower quartile (box). Whiskers extend to the smallest/largest value within 1.5 times the interquartile range. Dots represent extreme values.

Regarding the impact of scenario on confidence ratings, we see only a small difference between the scenarios. While a speaker uttering *believe* received numerically higher confidence ratings in the interrogation than in the briefing – going against our predictions – median confidence ratings for speakers uttering *know* and the bare assertion are similar across scenarios; see Figure 8.

The observations about the confidence ratings are supported by the outcome of the analysis, but not the scenario-specific variation. Averaged over scenarios, participants assigned certainty values below average (grand mean = 0.74 on the original scale) when assessing a speaker uttering *believe* ($\hat{\beta} = -0.83$, CrI: [-0.98, -0.68]) and slightly above average when assessing a speaker uttering the bare assertion ($\hat{\beta} = 0.16$, CrI: [0.05, 0.27]), see Table 6. Participants assigned the highest certainty values to a speaker uttering *know* (on average 0.67 above the grand mean/intercept on the log-odds scale). The results suggest that there is no evidence for an interaction of *believe* ($\hat{\beta} = -0.03$, CrI: [-0.09, 0.04]) or the bare assertion ($\hat{\beta} = -0.00$, CrI: [-0.07, 0.06]) and scenario in predicting speaker certainty.

Coefficient	posterior mean	Standard Error	l-95% CrI	u-95% CrI
Intercept (grand mean)	1.06	0.09	0.87	1.24
<i>believe</i>	-0.83	0.08	-0.98	-0.68
bare assertion	0.16	0.06	0.05	0.27
Scenario1	-0.07	0.04	-0.16	0.01
<i>believe</i> * Scenario1	-0.03	0.03	-0.09	0.04
bare assertion * Scenario1	-0.00	0.03	-0.07	0.06

Table 6: Population-level estimates of the Bayesian beta regression model on the log-odds scale with 95% credible intervals. The intercept is the grand mean and translates to 0.74 on the original scale. The categorical predictor formulation was sum-coded, see (4). The coefficient for *believe*: (-0.83) is the difference between intercept and *believe*, the coefficient for the bare assertion (0.16) is the difference between intercept and the bare assertion. The difference between the intercept and *know* can be calculated based on the coding for *know*, i.e. *believe*/bare assertion = -1: $-0.83 * (-1) + 0.16 * (-1) = 0.67$. The effect scenario is the change in log-odds for the briefing (-1 interrogation, 1 briefing). Slope coefficients whose 95% credible intervals do not include zero and are therefore treated as reliable effects are highlighted in bold.

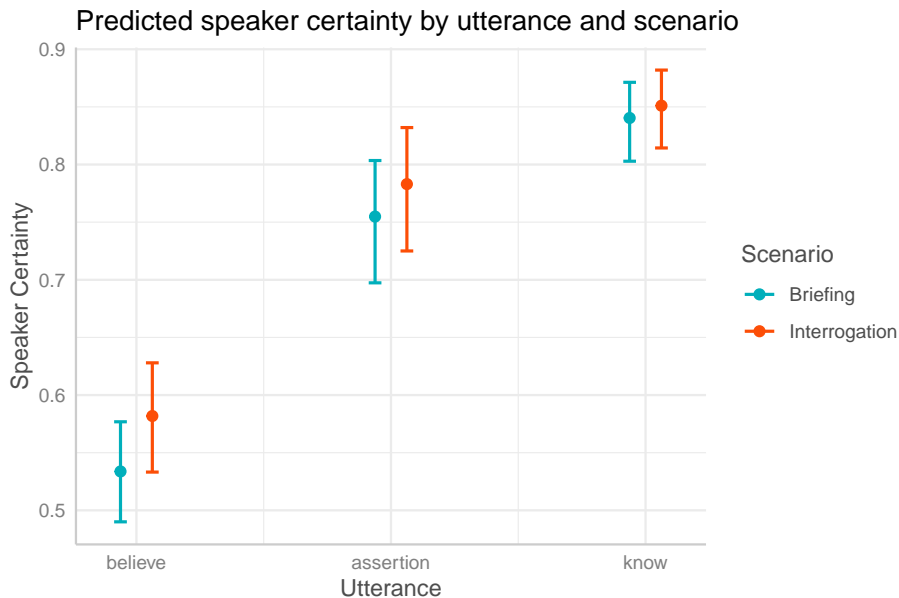


Figure 9: Predictions for the comprehension data. Log-odds were back-transformed to speaker certainty (y-axis). This figure shows the assigned speaker certainties by formulation and scenario (briefing in blue, interrogation in red) with error bars representing the 95% credible intervals.

We used the same method as in section 3.1.6 to illustrate how ratings vary across scenario and

formulation. Figure 10 shows the differences in assessed speaker confidence between the three formulations in the interrogation relative to the briefing. Given our model and our data we do not have evidence that there is an effect of scenario on the comprehension of either *know* or the bare assertion. There is only a weak tendency for *believe* such that participants would assign higher confidence ratings to a speaker uttering *believe* in the interrogation than in the briefing.

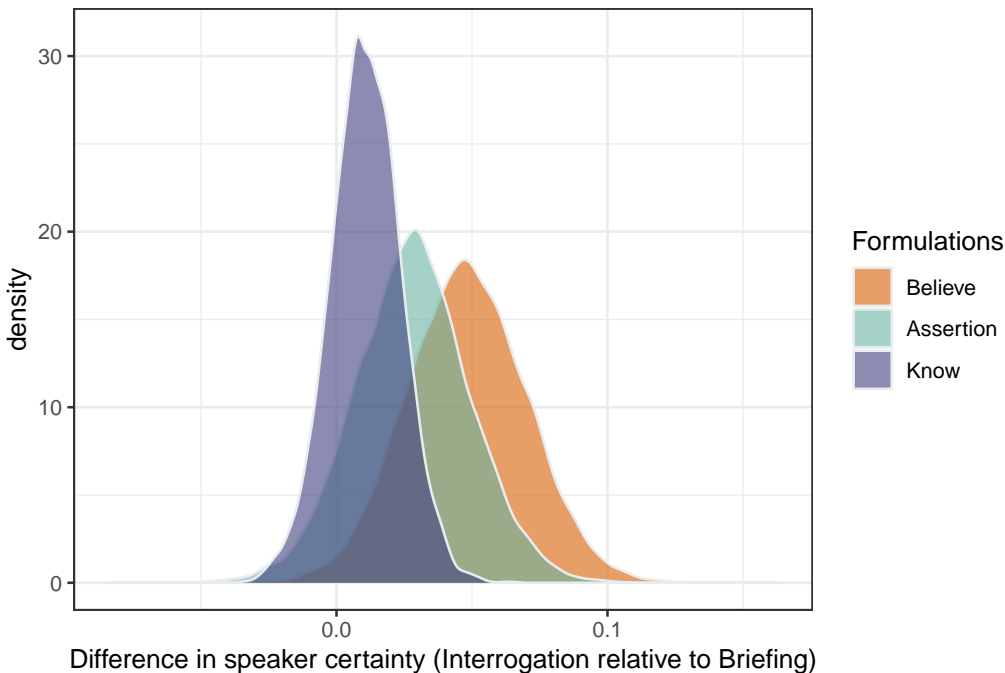


Figure 10: The histogram shows the differences in the assessment of speaker confidence in the interrogation relative to the briefing for *believe* (orange), bare assertion (green) and *know* (violet).

3.2.7 Discussion

In keeping with Experiment 1, Experiment 2 shows that participants assigned confidence ratings in same relative order: *believe* < bare assertion < *know*. Overall, there was no clear influence of the scenario manipulation on the confidence ratings. The most distinct difference was found for *believe*: participants would assign marginally lower certainties to speakers uttering *believe* in the briefing than in the interrogation which is at odds with our predictions. We cannot exclude the possibility that there were no differences in assigned confidence ratings.

Turning to the production of *know*, we could only find a marginal difference between scenarios but this effect runs in the opposite direction to that predicted: participants assigned marginally higher certainty values to a speaker uttering *know* in the interrogation than in the briefing. This

would mean that the participants/hearers of experiment 2 misunderstood the participants/speakers of experiment 1: Participants/hearers assigned higher speaker confidence values than the speakers' actual degrees of belief. However, we cannot be confident in the reliability of this effect, since the coefficient lies between -0.16 and 0.01 with 95% probability, a range which includes 0.

4 General Discussion

4.1 Which is the strongest formulation?

Our production and comprehension results suggest that *know* is epistemically the most powerful formulation. More specifically, speakers use *know*, the bare assertion and *believe* for content with successively lower evidentiality scores, and hearers likewise infer greater speaker confidence from the use of *know* than from the bare assertion, which in turn is associated with more confidence than the use of *believe*. Thus, one could interpret these results as evidence for an account similar to Mandelkern and Dorst (2022), according to which *believe* and *know* are associated with epistemic states such as weak belief and strong belief/knowledge respectively, and while the norm of assertion is typically weak belief these standards may shift depending on the requirements of the context. Indeed, the bare assertion was on average chosen for lower evidentiality ratings in the interrogation than in the briefing, suggesting that the context may be associated with different epistemic standards. We did not, however, find that the bare assertion was produced for a wider range of evidentiality ratings or inferred to convey a wider range of speaker certainty values than the other two formulations. One possibility is that this pattern is due to the setting of the experiment, which did not test contextual variation in form of varying epistemic standards discussed by Mandelkern and Dorst (2022). Hence, there might be greater variation in the production and comprehension of the bare assertion than could be revealed by our experimental set up. Furthermore, Mandelkern and Dorst (2022)'s account does not discuss any possible context effects that may influence the production and comprehension of *believe* and *know* which were tested in this study.

Overall, speakers who want to express some uncertainty about propositional content being true do so by producing *believe*. By doing so speakers may signal that they find this content should not be accepted to the common ground without further negotiation. In contrast, when wanting to express maximal certainty, speakers most frequently produce *know*. Proposing propositional content to be added to the common ground via the bare assertion may be an epistemically stronger proposal than using *believe* but a weaker one than using *know*. Hearers were found to infer the same relative ordering which could suggest that the epistemic strength with which speakers proposed content translated to the hearers. However, the way in which hearers would actually react towards these different formulations – i.e. whether p of *believe* p would be more frequently negotiated further

than if it was asserted, and so on – remains to be investigated.

4.2 Communicative goals

While epistemicity plays an essential role in the speakers' production choices, formulations may more or less suit the speakers' communicative goals. Our results suggest that speakers seem to utter *believe* differently depending on the scenario they are in. More specifically, speakers produced *believe* at numerically higher rates and higher evidentiality ratings in the briefing than in the interrogation. This suggests that speakers indeed use *believe* as a hedging device in cooperative settings. As was claimed previously, hedging may aid communication in cooperative situations such as the briefing in which interlocutors have different opinions but want to reach a shared common ground. Hedging may be used by speakers to signal their willingness to discuss their conveyed content further as well as their consideration for the beliefs of the other interlocutors.

While *know* was found to convey the highest degrees of belief and therefore may be described as the most powerful formulation, our results similarly suggests that speakers do not always choose to express their beliefs via *know* even if they are highly certain. More specifically, our results suggest that speakers produce *know* for lower degrees of belief in the interrogation than in the briefing such that speakers uttered *know* more readily – for lower degrees of belief and more frequently – in the interrogation. However, we acknowledge that these results are partly based on a weak interaction effect and warrant further investigation. Overall, it seems that presupposing factivity of new content can indeed be understood as an uncooperative discourse move which should be avoided in cooperative settings.

Although speakers adjust their production choices to the scenario they are in, we did not find evidence that hearers recover the basis for these production choices. The relevant coefficient for scenario lies between -0.16 and 0.01 with 95% probability, a range which includes 0. This can either mean that hearers do not distinguish between scenarios when assessing speaker confidence or that hearers assign higher speaker confidence in the interrogation opposed to the briefing. With regard to hedging, the first outcome would be at odds with findings that suggest that hearers are aware of hedging (e.g. Bonnefon & Villejoubert, 2006; Hosman, 1989). The second outcome would be at odds with our predictions that due to politeness constraints in the briefing speakers would use *believe* for higher evidentiality measures and hearers should be aware of this. With regards to the factivity of *know*, the first outcome may mean that, whereas *know* may be used more cautiously by speakers in cooperative settings, hearers infer that a speaker producing *know* must have a good reason to doing so, and therefore, assign high certainties across the board. However, it is less clear why hearers would infer higher certainties in the interrogation than in the briefing. One reason may

be that in the interrogation interrogators are interested in projecting higher confidence in front of the suspect which can feed into the perceived ratings of certainty.¹⁵

The fact that we did not find an effect of scenario in the predicted direction in the comprehension experiment may alternatively be due to our experimental design in which we investigated comprehension from a bystander point of view. In the production experiment, participants took over the role of the speaker who was talking to either a colleague or a suspect, whereas participants in the comprehension experiment were merely judging the speech of a second speaker who they did not directly interact with. Being a bystander might make it less relevant to reason about the speakers' strategic motivation than being the addressee. For example, a suspect in an investigation would be much more vigilant about what the interrogator is saying. Note that we chose the bystander point of view because we wanted to stick to the within-subject design of the production study. It would have been problematic to ask participants to take over the role of a suspect and pretend to be another person that might or might not have been involved in a criminal case. Participants might have not only end up reasoning about what the interrogator knows but also about their own confidence about what happened in the case. Additionally, we did not have a coherent cover story that would have made it plausible to ask the same participant to take over the role of a detective and a suspect in the same experiment.

4.3 A revised account

A speaker and hearer model that captures the production of *believe*, *know* and the bare assertion should aim to combine epistemic considerations and possible strategic usages for at least *believe* and possibly *know* that are not available for the bare assertion to the same extent. In the following we will sketch out a rough model for future investigations which has yet to be tested.

In order to capture the usage of the bare assertion, [Mandelkern and Dorst \(2022\)](#) proposed the beginnings of an account of epistemic posturing:

Epistemic posturing: In performing the speech act of asserting p , act as if you are absolutely certain of p .

Epistemic posturing only applies to a single speech act, such that there is no need to maintain the pretence of absolute certainty before or after the assertion. In this way the authors capture instances in which a speaker felicitously proclaims that they do not know that p but have either

¹⁵We thank an anonymous reviewer for this point.

previously or subsequently asserted p , such as in (13).¹⁶

- (13) a. Speaker1: What will John bring?
b. Speaker2: I just don't know.
c. Speaker1: Well, fine, but what do you think?
d. Speaker2: Hm, ok, let's see. He'll get Indian. That's his favourite, after all.
e. Speaker1: ?? How do you know he'll get Indian?

Example (13) similarly illustrates how the standards of assertion may be negotiated before and after acting in accordance with epistemic posturing. Here, the subsequent question 'How do you know?' is not appropriate, since it was previously established that Speaker2 does not know what John will bring. Example (14) illustrates this further. Here, the question 'Do you think that?' in (14b) feels redundant, because the norms of assertion according to Mandelkern and Dorst (2022) already require Speaker1 to believe what they say. In contrast, the question 'Do you know that?', is argued to be an effective way to negotiate the standards for assertion in this context.

- (14) Speaker1: John will bring Indian.
a. Speaker2: Do you know that?/ Are you sure?
b. Speaker2: ?? Do you think that?

The reason for why epistemic posturing is useful for communication is that it generally seems more informative for interlocutors to make a best guess than not to say anything at all in order to narrow down the space of alternatives. Thus, by forming their best guess, speakers are attempting to trade off optimally between informativity and accuracy. How this optimisation is performed, Mandelkern and Dorst (2022) argue, varies with the speakers' level of epistemic risk aversion and the demands of the context. A reason for why the bare assertion was used for on average lower evidentiality ratings in the interrogation than in the briefing may be due to a context-shift such that, for example, accuracy may have been less crucial than informativity in the interrogation. However, this line of reasoning remains speculative at this point.

Although epistemic posturing may account for the production of the bare assertion, it does not necessarily explain circumstances under which speakers would utter *believe* or *know* instead. To phrase it differently, it is not clear why speakers should ever utter *believe* or *know* if the bare assertion can be produced to convey a wide range of epistemic states. The first reason why speakers might produce *believe* and *know* over the bare assertion would be an epistemic one: the speaker

¹⁶This example may also serve as argument against Unger (1975)'s argument in support of the Knowledge Norm of Assertion.

may want to make explicit their epistemic state with regards to propositional content in order to signal how other interlocutors should treat that content. The second reason may relate to the cooperativity, or uncooperativity, of the communicative setting.

More specifically, based on our findings, speakers who want to explicitly communicate relatively low degrees of belief do so by producing *believe that p* instead of the bare assertion and in that way signal that *p* should not readily be accepted to the common ground and could instead be discussed further in the subsequent discourse. However, based on our findings, *believe that p* may similarly be used when speakers are fairly certain about *p*. Thus, a model capturing the production of all three formulations should account for the production of *believe* as a hedging device to come across as polite or, following more egocentric motives, to reduce accountability. In cooperative communication, such behaviour may be advantageous, since speakers may signal that they are willing to discuss content further and that they are open to other suggestions as to how they can reach their joint communicative goal. Therefore, in cooperative settings, informativity and accuracy may also be traded off with politeness, which signals willingness to cooperate or other more egocentric motives such as reducing accountability.

Our findings showed that speakers who want to communicate an epistemic state close to maximal certainty do so by producing *know that p*. However, in contrast to the bare assertion which can also be used to express maximal certainty, *know that p* was used slightly differently. Our results tentatively suggest that *know* was used more readily (for lower certainties and more frequently) in the uncooperative scenario. This may be interpreted as weak evidence for the claim that the production of *know* signals that *p* should already be part of the common ground. Hence, producing *know that p* when *p* is controversial may come across as uncooperative which is a disadvantage in a setting in which interlocutors aim to reach a joint communicative goal. Again, we acknowledge that this part of our results is weak and emphasise that more research is needed to test the predictions of this account further.

5 Conclusion

In this study, we tested the production and comprehension of the formulations *believe*/bare assertion/*know* in two contrasting scenarios. We found that *know* is an epistemically stronger formulation than the bare assertion. More specifically, speakers choose between these three formulations based on their epistemic state, and likewise, hearers infer the speakers' confidence based on their formulations. Regarding the formulations' conveyed degrees of belief, our findings suggest the following ranking: *believe* < bare assertion < *know*. While this implies that the norm of assertion is one of weak

belief whose standards may shift according to context, our findings also suggest that speakers may sometimes want to explicitly communicate their epistemic states via the production of *believe* and *know*. Furthermore, our findings shed more light on speakers' communicative goals when producing *believe* and *know*. Our findings are compatible with a view on which *believe* is used to hedge the assertive strength of statements in cooperative settings. We also found weak evidence for the claim that the production of *know* may sometimes be too powerful. More concretely, while *know* is epistemically the most powerful expression, its production may be dispreferred in a cooperative negotiation of controversial content. Whereas speakers and hearers agree on the relative epistemic ordering of the formulations, hearers do not appear consistently to take account of the speakers' communicative goals. As mentioned in the discussion, this may be due to the experimental design of the comprehension study where participants were in a bystander role, unlike the participants in the production experiment who were meant to be playing a more active role in the simulated interactions. Based on our findings we proposed a revised speaker-hearer account which aims to capture the production and comprehension of *believe*/bare assertion/*know* based on the formulations' conveyed degrees of belief, and the speakers' communicative goals.

Disclosure statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Beltrama, A. & Papafragou, A. (2023). Pragmatic violations affect social inferences about the speaker. *Glossa Psycholinguistics*, 2(1), 1–41. doi: Mhttp://dx.doi.org/10.5070/G601197
- Bonnefon, J.-F. & Villejoubert, G. (2006). Tactful or doubtful?: Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science*, 17(9), 747–751.
- Brown, P. & Levinson, S. (1978). Universals in language usage: politeness phenomena. In E. Goody (Ed.), *Questions and politeness. strategies in social interactions* (pp. 56–311). Cambridge University Press.
- Brown, P. & Levinson, S. (1987). *Politeness. some universals in language usage*. Cambridge: Cambridge University Press.
- Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the r package brms. *The R Journal*, 10(1), 395–411.
- Crismore, A. & Kopple, W. J. V. (1997). Hedges and readers: Effects on attitudes and learning. In *Hedging and discourse* (pp. 83–114). Berlin, Boston: De Gruyter.
- Franke, M., Dulcinati, G. & Pouscoulous, N. (2020). Strategies of deception: Under-informativity, uninformativity, and lies—misleading with different kinds of implicature. *Topics in Cognitive Science*, 12(2), 583–607.
- Fraser, B. (1975). Hedged performatives. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 187–210). New York: Academic Press.
- Granhag, P. A. & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind reading. *Psychology, Crime & Law*, 14, 189–200.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, vol. 3, speech acts* (pp. 41–58). New York: Academic Press.
- Hawthorne, J., Rothschild, D. & Spectre, L. (2016). Belief is weak. *Philos Stud*, 173, 1393–1404.
- Hoffman, M. D. & Gelman, A. (2013). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1), 1593–1623.
- Holmes, J. (1982). Expressing doubt and certainty in english. *RELC Journal, 1982-12*, 13(2), 9–28.
- Hosman, L. A. (1989). The evaluative consequences of hedges, hesitations, and intensifies powerful and powerless speech styles. *Human Communication Research*, 15(3), 383–406.
- Juanchich, M. & Sirota, M. (2015). A direct and comprehensive test of two postulates of politeness theory applied to uncertainty communication. *Judgment and Decision Making*, 10(3), 232–240.
- Kruschke, J. K., Aguinis, H. & Joo, H. (2012). The time has come: Bayesian methods for data

- analysis in the organizational sciences. *Organizational Research Methods*, 15(4), 722-752.
- Lackey, J. (2007). Norms of assertion. *Noûs*, 41, 594–626.
- Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2(4), 458–508.
- Lewis, D. (1979). Scorekeeping in a language game. In R. Bäuerle, U. Egli & A. von Stechow (Eds.), *Semantics from different points of view* (pp. 172–187). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Lorson, A., Cummins, C. & Rohde, H. (2021). Strategic use of (un)certainly expressions. *Frontiers in Communication*, 6.
- Luke, T. J. (2021). A meta-analytic review of experimental tests of the interrogation technique of hanns joachim scharff. *Applied Cognitive Psychology*, 35(2), 360-373.
- Mandelkern, M. & Dorst, K. (2022). Assertion is weak. *Philosophers' Imprint*. doi: M10.3998/phimp.1076
- Meibauer, J. (2014). *Lying at the semantics-pragmatics interface*. Berlin: Mouton de Gruyter.
- Moore, G. E. (1942). A reply to my critics. In P. Schilpp (Ed.), . Northwestern University.
- Ochs, E. (1976). The universality of conversational postulates. *Language in Society*, 5(1), 67–80. doi: Mhttp://www.jstor.org/stable/4166850
- R Core Team. (2020). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from [Mhttps://www.R-project.org/](https://www.R-project.org/)
- Schwarz, F. (2019). Presuppositions, projection and accommodation. In C. Cummins & N. Katsos (Eds.), *Oxford handbook of experimental semantics and pragmatics* (pp. 83–113). Oxford: Oxford University Press.
- Smithson, M. & Verkuilen, J. (2006). Better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11(1), 54–71.
- 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.
- Stalnaker, R. (1974). Pragmatic presuppositions. In R. Stalnaker (Ed.), *Context and content* (pp. 47–62). Oxford University Press.
- Stan Development Team. (2017). Stan modeling language users guide and reference manual, version 2.17.0 [Computer software manual]. Retrieved from [Mhttp://mc-stan.org/](http://mc-stan.org/)
- Sóskuthy, M. & Roettger, T. B. (2020, 09). When the tune shapes morphology: The origins of vocatives. *Journal of Language Evolution*, 5(2), 140–155.
- Unger, P. (1975). Ignorance: A case of skepticism.
- van der Sandt, R. A. (1992, 11). Presupposition Projection as Anaphora Resolution. *Journal of Semantics*, 9(4), 333-377.

von Fintel, K. (2008). What is presupposition accommodation again? *Philosophical Perspectives*, 22(1), 137–170.

Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.

6 Appendix I

6.1 Production Experiment: model output with uninformative priors

- by-expression intercepts: $\text{student}_t(3, 0, 2.5)$
- fixed effects: flat prior
- standard deviations: $\text{student}_t(3, 0, 2.5)$
- correlation matrix: LKJ(1)

Utterance	Coefficient	posterior mean	Est.Error	l-95% CrI	u-95% CrI	\hat{R}
believe	Intercept	1.20	0.22	0.77	1.62	1.00
bare assertion	Intercept	0.27	0.16	-0.06	0.59	1.00
believe	Evidentiality	-1.57	0.18	-1.94	-1.24	1.00
bare assertion	Evidentiality	-0.40	0.17	-0.72	-0.07	1.00
believe	Scenario	0.60	0.11	0.39	0.82	1.00
bare assertion	Scenario	0.28	0.11	0.06	0.50	1.00

Table 7: Population-level estimates of the categorical regression model in log-odds with the standard errors and 95% credible intervals. In the table the by-expression intercepts are listed first, then the estimates for the evidentiality effect followed by the estimates for the scenario effect. The effect scenario is the change in log-odds for the briefing (-1 interrogation, 1 briefing). \hat{R} is a convergence diagnostic which compares the between- and within-chain estimates. Values larger than 1 suggest that the chains have not mixed well.

6.2 Comprehension Experiment: model output with uninformative priors

- intercept: $\text{student}_t(3, 0, 2.5)$
- fixed effects: flat prior
- standard deviations: $\text{student}_t(3, 0, 2.5)$
- correlation matrix: LKJ(1)

Coefficient	posterior mean	Standard Error	l-95% CrI	u-95% CrI	\hat{R}
Intercept (grand mean)	1.06	0.10	0.87	1.24	1.00
Utterance I	-0.83	0.08	-0.99	-0.68	1.00
Utterance II	0.17	0.06	0.06	0.28	1.00
Scenario1	-0.07	0.04	-0.16	0.01	1.00

Table 8: Population-level estimates of the Bayesian beta regression model on the log-odds scale with 95% credible intervals. The categorical predictor formulation was sum-coded. The effect scenario is the change in log-odds for the briefing (-1 interrogation, 1 briefing).

7 Appendix II

7.1 Production Experiment: model output with order effect

Utterance	Coefficient	posterior mean	Est.Error	l-95% CrI	u-95% CrI	\hat{R}
believe	Intercept	1.18	0.26	0.68	1.70	1.00
bare assertion	Intercept	0.30	0.23	-0.15	0.73	1.00
believe	Evidentiality	-1.51	0.17	-1.86	-1.19	1.00
bare assertion	Evidentiality	-0.36	0.16	-0.67	-0.04	1.00
believe	Scenario	0.59	0.11	0.39	0.80	1.00
bare assertion	Scenario	0.27	0.11	0.06	0.49	1.00
believe	Order	0.00	0.03	-0.05	0.06	1.00
bare assertion	Order	-0.01	0.03	-0.06	0.05	1.00

Table 9: Population-level estimates of the categorical regression model in log-odds with the standard errors and 95% credible intervals. In the table the by-expression intercepts are listed first, then the estimates for the evidentiality effect followed by the estimates for the scenario effect. The effect scenario is the change in log-odds for the briefing (-1 interrogation, 1 briefing). \hat{R} is a convergence diagnostic which compares the between- and within-chain estimates. Values larger than 1 suggest that the chains have not mixed well.

7.2 Production Experiment: model without interaction

Formulation	Coefficient	posterior mean	Est.Error	l-95% CrI	u-95% CrI	\hat{R}
believe	Intercept	1.18	0.21	0.77	1.60	1.00
bare assertion	Intercept	0.26	0.16	-0.07	0.57	1.00
believe	Evidentiality	-1.51	0.17	-1.85	-1.19	1.00
bare assertion	Evidentiality	-0.36	0.16	-0.67	-0.04	1.00
believe	Scenario	0.59	0.11	0.39	0.81	1.00
bare assertion	Scenario	0.27	0.11	0.05	0.48	1.00

Table 10: Population-level estimates of the categorical regression model in log-odds with the standard errors and 95% credible intervals. In the table the by-expression intercepts are listed first, then the estimates for the evidentiality effect followed by the estimates for the scenario effect. The effect scenario is the change in log-odds for the briefing (-1 interrogation, 1 briefing). Slope coefficients whose 95% confidence intervals do not include zero and are therefore treated as reliable effects are highlighted in bold. \hat{R} is a convergence diagnostic which compares the between- and within-chain estimates. Values larger than 1 suggest that the chains have not mixed well.

7.3 Comprehension Experiment: model output with order effect

Coefficient	posterior mean	Standard Error	l-95% CrI	u-95% CrI	\hat{R}
Intercept (grand mean)	1.07	0.10	0.86	1.27	1.00
Utterance I	-0.83	0.08	-0.98	-0.68	1.00
Utterance II	0.16	0.06	0.05	0.28	1.00
Scenario1	-0.07	0.04	-0.16	0.01	1.00
Order	-0.00	0.01	-0.02	0.01	1.00

Table 11: Population-level estimates of the Bayesian beta regression model on the log-odds scale with 95% credible intervals. The categorical predictor formulation was sum-coded, see (4).

7.4 Comprehension Experiment: model without interaction

Coefficient	posterior mean	Standard Error	l-95% CrI	u-95% CrI	\hat{R}
Intercept (grand mean)	1.05	0.09	0.86	1.24	1.00
Utterance I	-0.82	0.08	-0.97	-0.68	1.00
Utterance II	0.16	0.06	0.05	0.28	1.00
Scenario1	-0.07	0.04	-0.16	0.01	1.00

Table 12: Population-level estimates of the Bayesian beta regression model on the log-odds scale with 95% credible intervals. The intercept is the grand mean and translates to 0.74 on the original scale. The categorical predictor formulation was sum-coded, see (4). The coefficient for Utterance I (-0.82) is the difference between intercept and *believe*, the coefficient for Utterance II (0.16) is the difference between intercept and the bare assertion. The difference between the intercept and *know* can be calculated based on the coding for *know*, i.e. $\text{UtteranceI/UtteranceII} = -1: -0.82 * (-1) + 0.16 * (-1) = 0.66$. The effect scenario is the change in log-odds for the briefing (-1 interrogation, 1 briefing).

8 Appendix III

8.1 Production Experiment: sample experimental items

Did Emily Brown have any financial problems?



Evidence:

Statement of Emily Brown's friend 20/12/2019
Name: Greta Lawson
Born: 01.02.1982
Issue: Emily Brown's financial situation

Campbell: "Did Emily Brown mention anything about her financial situation?"

Greta Lawson: "She mentioned that she wishes to win in the lottery to buy that house in the south of France that she wanted for so long."

Figure 11: Experimental item: Briefing (suspect Emily Brown, weak evidence).

Did Emily Brown have any financial problems?



Evidence:

HARRIOT BANK

RETURN SERVICE REQUESTED

Emily Brown
2/1 South Bridge
EH1 1LL Edinburgh

Statement Ending 30/11/2019
Page 1 of 4

Managing Your Accounts

- Primary Branch: Edinburgh
- Phone Number: XXXXX XXXXXX
- Online Banking: HarriotBank.com
- Mailing Address: 2/1 South Bridge EH1 1LL, Edinburgh

Summary of Accounts

Account Type	Account Number	Ending Balance
HARRIOT RELATIONSHIP CHECKING	XXXXXXXX4101	£ -15,045

Primary Checking

Account Summary	Description	Amount
Date		
01/11/2019	Beginning Balance	£ -9,000
30/11/2019	Ending Balance	£ -15,045

Figure 12: Experimental item: Briefing (suspect Emily Brown, strong evidence).

Did you have any financial problems?



Evidence:

Statement of Emily Brown's friend 20/12/2019

Name: Greta Lawson

Born: 01.02.1982

Issue: Emily Brown's financial situation

Campbell: "Did Emily Brown mention anything about her financial situation?"


Greta Lawson: "She mentioned that she hopes to win in the lottery to buy that house in the south of France that she wanted for so long."

Figure 13: Experimental item: Interrogation (suspect Emily Brown, weak evidence).

Did you have any financial problems?



Evidence:

 HARRIOT BANK

RETURN SERVICE REQUESTED

Emily Brown
2/1 South Bridge
EH1 1LL Edinburgh

Statement Ending 30/11/2019
Page 1 of 4

Managing Your Accounts

- Primary Branch: Edinburgh
- Phone Number: XXXXX XXXXX
- Online Banking: HarriotBank.com
- Mailing Address: 2/1 South Bridge EH1 1L, Edinburgh

Summary of Accounts

Account Type	Account Number	Ending Balance
HARRIOT RELATIONSHIP CHECKING	XXXXXXXXX101	£ -15,045

Primary Checking

Date	Description	Amount
01/11/2019	Beginning Balance	£ 9,900
30/11/2019	Ending Balance	£ -15,045

Emily Brown's bank statement from November 2019

Figure 14: Experimental item: Interrogation (suspect Emily Brown, strong evidence).

9 Appendix IV

9.1 Comprehension Experiment: sample experimental items



Figure 15: Experimental item: Briefing (suspect Emily Brown/'know').

Did you have any financial problems?



Figure 16: Experimental item: Interrogation (suspect Emily Brown/‘believe’).

Were you ever in trouble with the police before?

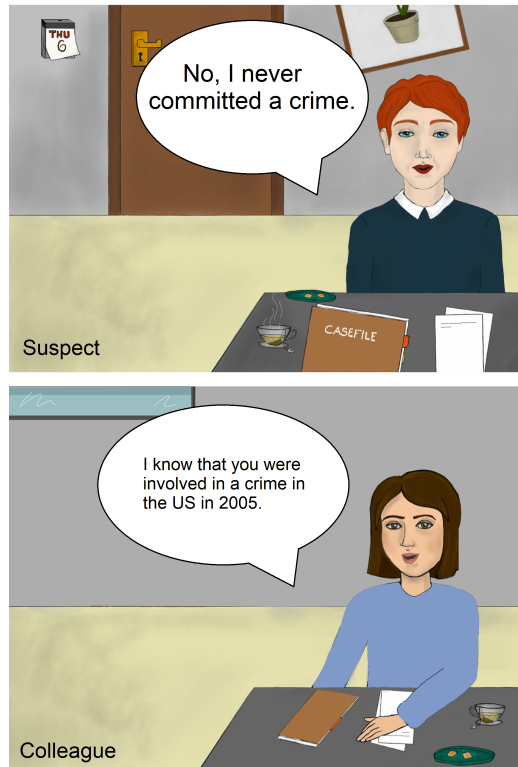


Figure 17: Experimental item: Interrogation (suspect Johanna Smith/'know').

9.2 Comprehension Experiment: sample control items



Figure 18: Control item/attention check: Briefing



Figure 19: Control item/attention check: Interrogation