This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
Choosing to presuppose:
Strategic uses of presupposition triggers

Alexandra Lorson
The University of Edinburgh

Department of Linguistics and English Language
Supervisors: Chris Cummins & Hannah Rohde
15th August 2022
Acknowledgements

Firstly, I would like to thank my brilliant PhD supervisors Dr. Hannah Rohde and Dr. Chris Cummins for their continuous support. Among many things, you taught me how to connect the loose ends of my ideas and tell the big-picture story, as well as the importance of getting into the nitty-gritty details. I thank you for always being there even while we were on different continents or confined to our homes during the pandemic. Your contribution made me a better, more nuanced researcher and writer and elevated the value of this thesis greatly. I also would like to thank the Scottish Graduate School for Arts & Humanities in conjunction with the Scottish Funding Council for their financial support of this thesis. A big thank you as well to Belle Yeung for being my research assistant. You were a great help and excellent in acting as the wonderful suspect Emily Brown during experiment 1.

I would like to thank the department of English Language and Linguistics, the Meaning and Grammar research group as well as my PhD pals for making my PhD experience one that I will always look back on fondly. The lunches on the 7th floor of DSB were a daily highlight; both as a cohort and later with the remaining ladies. I specifically thank my dearest PhD pals: my äpple Lisa the chin and GIF master (over-sharing is caring!); Mattia the best flatmate (fiattexia forever) and husband of all times; Sarah the greatest swim pal and baking pro (1000 Christmas cookies are not enough). Without you my time as a PhD student would have been grey, boring and sad.

Furthermore, I would like to thank Donald(us) for all the plants, runs/hikes, doughnuts, sticky toffee puddings and for providing me with a home in Scotland. I adored all of our adventures which peaked with exploring deserted Scottish golf courses during the pandemic (the best kind of golf courses really). On to new adventures in Birmingham! I would also like to thank my family. Mum, thank you for always having an open ear for all sorts of sorrows and always being a rock (your advice means the world). Dad, thank you for always lending a helping hand (our road trip to Edinburgh when I first moved here was a blast), or one of your German classics. Lauri, thank you for being able to understand and relate when no one else can. I also thank Etti and Lydi for simply
being lovely friends! A PhD is a marathon with academic and emotional setbacks and without these amazing people would not have been possible to finish.
Abstract

This PhD project investigates the discourse structuring and managing properties of presupposition triggers. Specifically, the thesis is a theoretical and experimental investigation of what motivates speakers to presuppose as opposed to assert content, and how this decision influences the course of the subsequent discourse. This thesis is divided into three parts: (I) the negotiation of new content that is introduced by presuppositions versus assertions [experiment 1], (II) the communication of degrees of beliefs via factive presupposition triggers in contrast to uncertainty expressions and the bare assertion [experiments 2–5], and (III) the signalling of parallel information via additive presupposition triggers [experiments 6–7].

The first part of the thesis teases apart the distinction between presupposed/asserted and at-issue/not at-issue content by investigating whether (i) presupposed at-issue content is inaccessible or less accessible for interlocutors than asserted at-issue content, and (ii) in what way interlocutors address presupposed content in the subsequent discussion if the presupposed content is indeed less accessible. The experimental results suggest that, when material is relevant to the discourse question, it can be challenged directly by a subsequent speaker whether it is formally asserted or presupposed. However, expressing relevant material through presupposition rather than assertion may reduce the frequency of such challenges. Thus, a speaker-hearer model would have to include interlocutors’ expectations both about information packaging and about the overall discourse topic, in order to determine which content material may be discussed further.

The second part of the thesis focuses on the factive presupposition trigger know and its discourse structuring properties. For this purpose, experiments [2–3] investigated the speakers’ motivations in choosing between uncertainty expressions such as think, believe or the factive verb know in cooperative versus uncooperative scenarios. The results of both experiments suggest that speakers’ choice of formulation is influenced by (i) how likely they estimate an event to be and (ii) strategic considerations relating to the communicative context in which they are working. More specifically, speakers uttered know more frequently and for lower degrees of belief in uncooperative settings.
than in cooperative settings. Experiments [4–5] explored whether the strategic use of *know* in the uncooperative setting has to do with its discourse structuring properties as a presupposition trigger: By presupposing content speakers assume or act as if the conveyed information was already shared knowledge and not up for debate. Thus, hearers might be more inclined to accept and accommodate presupposed content than asserted content. For this purpose, speakers’ production choices and hearers’ interpretations of *know* versus the bare assertion were assessed. The results suggest that presupposing might have an advantage over asserting when speakers want to avoid further discussion of a topic. However, there was no evidence that hearers consider the speakers’ strategies when assessing their degrees of belief, which may mean that speakers can employ these communicative strategies successfully, though this point requires further investigation.

The third part of the thesis investigates the discourse managing properties of additive particles such as *too*, which are argued to presuppose a propositional alternative. If a suitable antecedent for the presupposition is present in the preceding dialogue, the production of *too* has been argued to be obligatory. Experiments [6–7] test (i) the potential obligatoriness and discursive functions of additives by manipulating the antecedents’ salience focusing on the factors Similarity and Turn Distance; and (ii) whether speakers would purposely violate the obligatoriness of *too* and with that avoid signalling similarity between what they and the antecedent speaker said in order to socially distance themselves from the antecedent speaker. Overall, the results of experiments [6–7] suggest that while the production of additives seems to depend on the antecedent’s salience, additive production was not as frequent as expected if additives were indeed obligatory. More specifically, participants were found to utter additives more frequently when their utterance’s content matched the content of a previously formulated utterance and when the matching utterance directly preceded their utterance. Furthermore, results of experiment [7] suggest that speakers deliberately drop the use of additives when wanting to diverge from an impolite speaker.

This project sheds more light on the speakers’ motivation to presuppose rather than assert content and on the way their production choices influence the preceding dialogue and the hearers’ interpretation. Overall, the results suggest that the speakers’ choice to either assert or presuppose content depends on (i) what content is part of the common ground, and (ii) speakers’ communicative goals: specifically in whether or not they are being maximally informative and cooperative. The thesis draws parallels between the discursive properties of different kinds of presupposition triggers while at the same time also highlighting possible differences between them, in particular between factive and additive presupposition triggers. By demonstrating the strategic use of presupposition triggers, this thesis informs theories on presupposition accommodation, the common ground, and communicative strategies in cooperative and uncooperative contexts.
Contents

Acknowledgements i
Abstract ii

Introduction 2

1 The difference between presupposing and asserting 2
   1.1 Presuppositions ........................................... 2
   1.2 Addressability of content ............................... 6
   1.3 Discourse structuring ................................... 7
   1.4 Discourse managing ..................................... 9

I Hiding lies in presuppositions? 11

2 When objecting to presupposed content comes easily 12
   2.1 Abstract .................................................... 12
   2.2 Introduction ............................................... 12
   2.3 Background ............................................... 13
       2.3.1 Strategic use of presuppositions .................. 13
       2.3.2 The role of at-issueness ............................ 14
   2.4 Experiment [1] ............................................. 16
       2.4.1 Materials and design ................................ 17
       2.4.2 Participants .......................................... 17
       2.4.3 Results ............................................... 17
   2.5 Discussion .................................................. 18
   2.6 Conclusion .................................................. 21
   2.7 Bayesian Analysis ........................................ 22
II Discourse Structuring

3 The difference between knowing and believing
   3.1 Abstract ........................................... 28
   3.2 Introduction ........................................ 28
   3.3 Previous work on (un)certainty expressions ................. 30
      3.3.1 Threshold semantics ................................ 30
      3.3.2 Communicative strategies .......................... 32
   3.4 The present study ...................................... 36
      3.4.1 Experiment 1 [2]: seven (un)certainty expressions ... 38
      3.4.2 Experiment 2 [3]: know versus believe ............... 47
   3.5 General Discussion ..................................... 51
   3.6 Conclusion ........................................... 56
   3.7 What’s next? .......................................... 57

4 Strategic reasoning when formulating and comprehending knowledge ascriptions 58
   4.1 Abstract ........................................... 58
   4.2 Introduction ........................................... 59
   4.3 Epistemic Strength ...................................... 60
      4.3.1 The Knowledge Norm of Assertion and Contextualism ... 60
      4.3.2 Empirical Findings .................................. 63
   4.4 Are factives a communicative strategy? ....................... 64
      4.4.1 The Presupposition Strength Account .................. 64
   4.5 The present study ...................................... 69
      4.5.1 Experiment 1 [4]: Production .......................... 70
      4.5.2 Experiment 2 [5]: Comprehension task ................. 86
   4.6 Discussion ............................................ 96
      4.6.1 Epistemic Accounts .................................. 96
      4.6.2 Presupposition Strength Account ....................... 98
      4.6.3 A revised account .................................. 99
   4.7 Conclusion ............................................ 101
   4.8 What’s next? .......................................... 101
Appendix B

7.2 Experiments [2–3]: Sample experimental items ........................................ x

Appendix C

7.4 Production experiment [4] ........................................................................ xx
7.4.1 Production experiment [4]: Model output with uninformative priors ........ xx
7.4.2 Production experiment [4]: Model output with order effect .................. xx
7.4.3 Production experiment [4]: Model predictions with interaction ............ xxii
7.4.4 Production experiment [4]: Sample experimental items ..................... xxiii

Appendix D

7.6 Experiment I [6] .................................................................................. xxxii
7.6.1 Experiment I [6]: Model output with uninformative priors .............. xxxii
7.6.2 Experiment I [6]: Model output with order effect ............................ xxxiii

List of tables and figures

xl

List of Tables

xl

List of Figures

xlii
Introduction
Chapter 1

The difference between presupposing and asserting

1.1 Presuppositions

When communicating, speakers rely on assumptions about what their interlocutors already know. Thus, speakers presuppose knowledge which is usually crucial for understanding each other. For example, writing a thesis in English presupposes that the audience will be able to understand English and can read. This thesis will focus on those presuppositions that are triggered by lexical items – so called presupposition triggers – such as stop, know and too – and investigates possible motives that speakers may have to presuppose rather than assert content.

Presupposition triggers signal a variety of distinct presuppositions. The change of state verb stop for instance triggers a presupposition in the form of ‘X used to V’, where X may denote an individual and V a habit/occupation: by uttering (1), the speaker presupposes that Mattia used to work for Starbucks.

(1) Mattia stopped working for Starbucks.

Factive presupposition triggers such as know and discover signal a factive presupposition in the form of ‘it is indeed true that Y’, where Y stands for the content of the factive’s complement. For example, in (2) the speaker presupposes that the joiner indeed planned to come over this afternoon.

(2) Donald knows that the joiner planned to come over this afternoon.

Additives such as too and also trigger a propositional alternative in the form of ‘there is a salient individual X (other than the individual of the sentence containing the additive) for whom the
property P holds true’. Applied to a concrete case, in (3), the speaker presupposes that someone else other than Etka – salient in the context – ordered coffee.

(3) Etka ordered coffee, too

This section will introduce some features that were used to distinguish presuppositions from other aspects of meaning such as assertions and aims to relate concepts (e.g. common ground, at-issueness) that are crucial to this thesis with each other.

(1) presuppositions are often described as being already shared knowledge between interlocutors (van der Sandt, 1992), ‘taken for granted’ (Chierchia & McConnell-Ginet, 2000) and part of the interlocutors’ common ground, whereby the common ground is said to contain information that is assumed to be true by the interlocutors of a conversation (Stalnaker, 1974). In contrast, assertions most commonly extend the common ground, unless they are directly objected to by interlocutors. It is therefore claimed that uttering a presupposition faithfully is contextually restricted such that the content of the presupposition should already be part of the common ground.

Furthermore, (2) presuppositions are often described as not being the main point of the utterance (Abbott, 2000) but instead argued to denote background information (Levinson, 1983). Another way of phrasing this is to say that presuppositions generally convey content that is not at-issue (Potts, 2005) with respect to the Question Under Discussion (QUD), whereby the QUD corresponds to the current discourse topic, following Roberts (1996). Not at-issue content then contrasts with content that is focused and generally addresses the QUD. Features (1) and (2) are connected such that already shared knowledge between interlocutors is mostly likely not useful to answer the current QUD.

Assuming presuppositions convey already known, not at-issue content that is not meant to be discussed further, (3) presuppositions are often claimed to require additional effort to be addressed in the subsequent discourse opposed to assertions. For example, if a speaker responds with (4b) to (4a), they are most often taken to deny the assertion (that Speaker1 visited their brother) rather than the presupposition (that Speaker1 has a brother).

(4) a. Speaker1: I visited my brother.
   b. Speaker2: That’s not true!/ No!

The ‘Hey, wait a minute’ test, proposed by von Fintel (2008) following Shannon (1976), is a diagnostic for presuppositions that utilizes this observation. Circumlocutions such as ‘Hey wait a minute...’,

\footnote{Focus is understood as the relation between prosodic prominence–often realised by a nuclear pitch–and pragmatic/semantic effects, following (Jackendoff, 1972).}
as in (5a), are claimed to be necessary in order to deny the presupposition, whereas they constitute a dispreferred means to deny asserted content, as shown in (5b).

(5)  a. Hey wait a minute, you don’t have a brother!
    b. ?Hey wait a minute, you didn’t visit your brother!

Lastly (4), another feature that has been used to identify whether content is presuppositional is to see whether that content projects out of entailment-cancelling environments such as negation. Unlike other aspects of meaning, presuppositions are said to project from under the scope of these operators (here, negation): (6) conveys the same presupposition as (2).

(6) Donald didn’t know that the joiner planned to come over this afternoon.

In contrast, taking entailments as an example, when (7a) is negated in (7b), the entailment that the window is broken does no longer hold.

(7)  a. Sarah didn’t break the window.
    b. Sarah didn’t break the window.
    c. #The window is broken.

To sum up, presuppositions usually convey content that is part of the common ground which is generally not very useful to settle open questions and therefore usually not at-issue which may contribute to presupposed content being less addressable. Being not about the main-point of an utterance may similarly contribute to the observed projection behaviour of presuppositions which seemingly escape entailment-cancelling operators.\(^2\)

However, the situation is more complicated than initially suggested. (1) presuppositions can be used to introduce new content. In such a case, following Lewis (1979) and von Fintel (2008), presupposition accommodation is argued to take place: the speaker acts as though a presupposition is already part of the common ground, and the hearer responds by adjusting their world-view to incorporate that presupposition. Realistically speaking, interlocutors often do not know what information they actually share, and may therefore rely on accommodation to take place.

\(^2\)Presuppositions may be more difficult to address than asserted content and escape entailment-cancelling operators. However, presuppositions are not impossible to address and also do not escape all forms of negation, see (1), where the presupposition that Kati used to smoke is cancelled via meta-linguistic negation.

(1) Kati didn’t stop smoking—she never started in the first place.
Furthermore, (2) novel presupposed content can be at-issue, as a speaker could use it to answer the QUD. Consider the exchange (8a)-(8b). Assuming (8a) is the QUD, (8b) answers indirectly by presupposing that the speaker had smoked. This material is clearly at-issue, as it does indeed answer the QUD, which the non-presupposed content (that the speaker does not currently smoke) does not.

(8)  
   a. Have you ever smoked?
   b. I quit smoking last year.

Being possibly new and at-issue content, then influences (3) the extent to which presupposed content is addressable. In fact, findings by Potts (2008) suggest that ‘Hey, wait a minute’ style circumlocutions may not exclusively be necessary to object to presupposed content but all types of not at-issue content. This would make the ‘Hey, wait a minute’ test insufficient to illustrate whether presupposed content is more difficult to address than asserted content.

Similarly, with regards to (4) it has been recently challenged that projection behaviour is something unique to presuppositions. Instead, Simons, Tonhauser, Beaver and Roberts (2010) and later Simons, Tonhauser, Beaver and Roberts (2016) propose that not only presuppositions but all implications that survive entailment-cancelling embeddings can project. This includes conventionally triggered implicatures like non-restrictive relative clauses which ought to project but are not presupposed (Potts, 2005). Simons et al. (2010, 2016) argue that embedded implications potentially project as long as they are not-at-issue relative to the discourse topic, i.e. the QUD, based on the assumption that only at-issue content is targeted by entailment-cancelling operators. In agreement with this account, the extent to which interlocutors find presuppositions projective was found to be influenced by prosodic cues that may render their conveyed content at-issue (Cummins & Rohde, 2015; Tonhauser, 2016). The more at-issue presupposed content was, the less projective such content was found to be.

The discussion above illustrates that it would be difficult to base a definition of presuppositions on any of the introduced features. Instead, in what follows, the terms ‘presupposition’/‘presupposed content’ will be used to denote the abstract form of the content that a given presupposition trigger conventionally signals (e.g. ‘X used to V’/‘it is indeed true that Y’). When using these terms, it will not automatically be assumed that the content conveyed by the presupposition is necessarily part of the common ground, not at-issue, backgrounded and projective. In fact, most parts of this thesis focus on speakers’ motivation to introduce new content that is at-issue via presupposition rather than assertion, similar to (8).
The reminder of the chapter will introduce the reader to the overall research topic that aims to answer this thesis’ research question repeated below:

RQ What motivates speakers to presuppose rather than assert content?

I will explain in which way the different parts of the thesis tackle this question and how they relate to each other. Furthermore, for each part, I will highlight the specific theoretical background against which our findings are discussed. Since this is a thesis-by-publication such that Chapters 2–5 each represent a collaborative paper, the pronoun ‘we’ is used to reflect this collaborative effort when applicable. While the experiments of each paper are numbered independently from the overall thesis, I will represent the overall numbering in the context of this thesis in square brackets. In Chapters 2–5 these numbers will appear after the paper internal numbering, e.g. ‘experiment 1 [3]’, ‘1’ referring to the paper internal and ‘3’ to the thesis numbering.

1.2 Addressability of content

Part I of the thesis focuses on the addressability of presupposed content (discussed as feature (3) in the above discussion), which may be distinct from the addressability of asserted content, by investigating a wide range of presupposition triggers: stop, know, regret, discover, return, only, to be annoyed, to be happy. The addressability of content may have to do with (i) its status of being either presupposed or asserted content, or (ii) the extent to which it is at issue towards the Question Under Discussion (QUD; based on Roberts (1996)). Keeping at-issueness constant by making both presupposed and asserted content relevant to the QUD, the aim of the first part of this thesis is to answer the following two questions:

(A) Is presupposed at-issue content more readily accepted by interlocutors than asserted at-issue content?

(B) In what way do interlocutors address presupposed at-issue content in the subsequent discussion?

The results of experiment [1] of this thesis suggest that (A) hearers are more likely to accept false presupposed content that is at-issue than false asserted content that is at-issue. However, we only found weak evidence. Moreover, (B) there was no evidence that addressing presupposed at-issue content was more effortful than addressing asserted at-issue content. Hence, we cannot exclude the

\[3\] My contribution in all cases was a first-author role: the development of the research question alongside my co-authors, the design and implementation of the methods, the full analysis, and the writing of the first draft and the catalyst in the revision of the papers.
possibility that presuppositional status plays a role in whether interlocutors pick up on content. However, ‘Hey wait a minute’-circumlocutions seem not to be necessary when speakers aim to address content that is at-issue. From a speaker’s perspective this may mean that speakers can influence what content is being discussed further. Specifically, by presupposing content speakers can reduce the chance that hearers pick up on that content.

1.3 Discourse structuring

Whereas Part I treats presupposition triggers as a generalized class, Part II and Part III focus on the production of specific presupposition triggers and the potential discursive role they play. Part II investigates the discourse structuring properties of the factive presupposition trigger know in contrast to (un)certainty expressions such as believe and the bare assertion. Since we did not find conclusive results in experiment [1] we investigated whether presupposed content is more readily accepted than asserted content further but from a speaker perspective. Part II tested the following research question:

(C) Do speakers utilize presupposing as a discourse structuring tool?

We understand discourse structuring as a process by which interlocutors may influence future discourse moves, e.g. what content is being picked up in the subsequent discourse. More specifically, Part II investigates whether hearers’ tendency to accommodate presupposed content may be exploited by speakers who aim to maximise the chances of the hearer accepting the proposition without further discussion: by presupposing rather than asserting information, speakers would act as if (potentially controversial) material was already accepted as part of the common ground and signal that they do not consider it to need further discussion. In order to investigate this hypothesis, the experiments [2–5] of Part II focus on the contrast between cooperative and uncooperative communication. Uncooperative speakers would be more likely than cooperative speakers to sneak in potentially controversial content into the conversation. Cooperative speakers would instead choose to make the potentially controversial content available for discussion by asserting it.

With regard to the formulations under investigation, (un)certainty expressions such as think and believe convey the speakers’ degrees of belief about the truth of their sentential complement, similarly to know. For example, a speaker uttering know that p conveys higher confidence in the proposition p being true than a speaker uttering believe that p. Whereas the distinction between (un)certainty expressions and know regarding their conveyed degrees of belief seems clear, the distinction between bare assertion and know is not. Thus, we discuss our results on the production and comprehension of know and the bare assertion in the context of philosophical accounts such as the Knowledge
Norm of Assertion (Williamson, 2000) and Epistemic Contextualism (DeRose, 2002, 1992). Within this context it is disputed whether speakers have to meet higher epistemic requirements to felicitously utter the bare assertion or *know*. Our production and comprehension findings resulted in the following ranking for the formulations’ degrees of belief: *believe*<bare assertion>*<know*. That is to say, our results indicate that the felicitous production of *know* is associated with higher epistemic requirements than the production of the bare assertion. This result is surprising under accounts which claim that the felicitous production of the bare assertion requires the speaker to be maximally certain.

Apart from communicating potentially different degrees of belief, (un)certainty expressions, the bare assertion and *know* differ from each other regarding the way they affect the discourse. (Un)certainty expressions have been contrasted with the bare assertion in the literature due to their being used as hedging devices: Whereas speakers who assert p propose to add p to the common ground, (un)certainty expressions can be used to *downtone* (Holmes, 1982) or *hedge* (Lakoff, 1973; Fraser, 1975; Brown & Levinson, 1978) the assertive strength of p, yielding discourse effects such as politeness. Thus, we discuss our results on the production and understanding of *believe* in contrast to the bare assertion in the context of politeness theory and corroborate the claim that *believe* is used as a hedging device.

The contrast between the discourse effects of *know* and the bare assertion was the main focus of Part II, see research question (C). Our production results suggest that speakers seem to presuppose content more readily – i.e. more frequently and for lower degrees of belief – when speaking to a potentially uncooperative interlocutor who might want to deny that content as opposed to a cooperative interlocutor. In contrast, the production of the bare assertion did not suggest such a pattern. Moreover, our comprehension results did not find a strong difference between uncooperative and cooperative communication. A slight trend suggests that hearers may assign higher degrees of belief to a speaker uttering any of the three formulations in an uncooperative setting. Both possibilities – no difference between cooperative and uncooperative communication or that hearers assign higher degrees to a potential uncooperative interlocutor – would suggest that speakers may be successful in their strategic use of *know*. However this has to be investigated further. With regard to the overall research question of this thesis, Part II suggests that one of the motives for speakers to presuppose content by uttering *know* is to structure the discourse to their liking.
1.4 Discourse managing

Part III investigates the discourse managing properties of additive presupposition triggers such as *too* whose use has been argued to be (i) only felicitous when there is a suitable antecedent in the preceding context, and (ii) obligatory if there is a suitable antecedent in the preceding context. The following research question was tested:

(D) Do speakers utilize presupposing as a discourse managing tool?

We understand discourse managing as a process by which interlocutors indicate the state of the common ground by referring back to content of the common ground. Part III tested whether speakers use additives simply because they are obliged to do so, or whether discourse and social factors influence the speakers’ production choices. Three factors were tested: Similarity, Turn Distance and Politeness. Whereas research findings suggest that Similarity influences the production of additives, Turn Distance has to our knowledge not been empirically investigated yet. Instead, empirical research has mostly focused on contexts in which the antecedent immediately precedes or is part of the host sentence. Furthermore, of interest was whether the production of additives can be seen as a means by which the speaker can socially converge with their interlocutor. Thus, the influence of Politeness on additive production was tested to see whether intentionally omitting additives in obligatory environments can be seen as an attempt by the speaker to diverge from an impolite antecedent speaker. For this purpose, two production experiments [6–7] were conducted. For experiment [6] participants were instructed to formulate their own responses to see whether participants use additives at all in the experimental setting and which kinds of additives they prefer. For experiment [7] a forced-choice paradigm was carried out in which participants were asked to choose between utterances with or without *too*.

The results of experiments [6–7] suggest that the production of additives is influenced by discourse factors: while speakers seem to use additives more frequently when there is a perfectly matching antecedent in the immediate context as predicted by accounts positing the obligatoriness of additives, speakers also used additives for partially matching antecedents, or antecedents that were uttered several turns previously. Moreover, the results of experiment [7] indicate that speakers use additives more frequently when speaking after a neutral antecedent speaker and omit additives more frequently when speaking after an impolite antecedent speaker. This suggests that the production of additives cannot solely be explained by additives being obligatory. Instead, I argue that additives play an important role in the participants’ grounding process which encompasses their discourse managing (D) and presuppositional properties: by using additives, speakers (i) refer back to content in the common ground, (ii) acknowledge parallel relations between content of the common ground
and their utterance, and (iii) signal that they kept track of what information is part of the common ground. All of these properties are relevant to grounding – a process by which interlocutors collect and coordinate knowledge (Clark, 1990). The use of additives can thus be understood as a cooperative participation in the grounding process whereby speakers converge with their interlocutors.

Seeing additives as a part of the interlocutors’ grounding process seems to capture our production data: whereby Turn Distance may explain cases when speakers unintentionally omit additives, Politeness may explain cases when speakers intentionally omit additives. More specifically, (i) not uttering additives may happen more frequently if the turn with the suitable antecedent happened a while back, since keeping track of what information is already in the common ground may become more difficult the more conversational turns pass, or the more time elapses; (ii) disengaging from the grounding process may also be due to a speaker’s attempt to diverge from the antecedent speaker. On top of that, (iii) speakers might not always deem it necessary to explicitly signal their involvement in grounding via additive use.

With regard to the overall research question of this thesis, Part III suggests that one of the motives for speakers to presuppose content by uttering additives such as too is to participate actively in the interlocutors’ grounding process. Moreover, the observed differences between speaking after a neutral antecedent speaker versus an impolite antecedent speaker may mean that there are contextual factors other than Similarity and Turn Distance influencing additive production that have yet to be investigated. In the Discussion, I will expand on the differences between presupposition triggers investigated in Parts II and III and discuss methodological shortcomings as well as the potential relevance of this thesis in general for other linguistic phenomena at the pragmatic/semantic interface.
Part I

Hiding lies in presuppositions?
Chapter 2

When objecting to presupposed content comes easily

2.1 Abstract

New content can be introduced into dialogue via presupposition as well as by assertion, but on traditional accounts presupposed information is expected to be less addressable in the subsequent dialogue. An alternative approach is to argue that addressability is more closely connected to whether content is at-issue with respect to the current Question Under Discussion. This paper investigates which of these factors is dominant. We report the results of a dialogue-based experiment designed to test whether and how false at-issue content is responded to in an ongoing discourse, and whether this is affected by its status as asserted or presupposed. Our findings suggest that when material is at-issue it can be challenged directly, independently of whether it is presupposed or asserted. However, relevant information introduced by a presupposition was found to be more likely to escape the participants’ attention.

2.2 Introduction

Speakers in dialogue can introduce new information in the form of presupposition: that is, by presenting it as though it were already part of the common ground. In the case of (9), the expression *my car* carries an existential presupposition to the effect that the speaker has a car, but (9) can nevertheless be uttered in a context in which the hearer does not already know this.

(9) Sorry I’m late: my car broke down.

Similarly, in cases such as (10) and (11), the presence of the expressions *quit* and *be happy* that
gives rise to presuppositions that might not already be known to the hearer, namely that John used
to smoke and that Mary’s boss is away.

(10) John is anxious. He quit smoking.
(11) Mary is happy that her boss is away.

This paper focuses on the status of new content that has been introduced via a presupposition. The
study we present uses an interactive dialogue paradigm to probe whether and how such content
is addressed as a discourse proceeds. The goal is to better understand how a speaker’s choice of
information packaging strategy within an individual utterance, considered alongside the active Quest-
ion Under Discussion across the broader discourse context, influences an interlocutor’s subsequent
discourse continuations and their ability to take up particular content.

2.3 Background

2.3.1 Strategic use of presuppositions

We can identify the meanings mentioned above (that John quit smoking, and that Mary’s boss is
away) as presuppositions of (10) and (11) on the basis of their ability to project from under the
scope of operators such as negation: (12) conveys the same presupposition as (10), and (13) as (11).

(12) John is anxious. He didn’t quit smoking.
(13) Mary isn’t happy that her boss is away.

Following Lewis (1979) and von Fintel (2008), the utterance of (9)-(13) to a hearer who lacks
the shared knowledge is argued to involve the exploitation of accommodation: the speaker acts
as though a presupposition is already part of the common ground, and the hearer responds by
adjusting their world-view, or situation model, to incorporate that presupposition. However, this
relies on the assumption that the presupposition is one that the hearer is willing to entertain (or at
least to ignore; see Glanzberg (2005)) rather than one that the hearer refuses to accept or wishes
to challenge. In normal cooperative conversation this assumption seems generally to be satisfied,
but it’s easy to find cases in which it is violated by a speaker deliberately introducing potentially
controversial material in the form of a presupposition, as in (14).\footnote{https://ottawacitizen.com/opinion/columnists/cohen-what-everybody-knows-about-america, retrieved 30 May 2019.}

(14) Everybody knows that Brett Kavanaugh’s confirmation was a farce.
Why, then, would a speaker choose to package information in the form of a presupposition rather than as a regular assertion? In the cooperative cases where the information is relevant but uncontentious, we could see this as arising partly from efficiency considerations – an utterance such as (9), (10) or (11) is more concise than the corresponding version in which the presupposed content is directly asserted (‘I have a car and it broke down’, etc.). But independent of efficiency, speakers might also select particular ways of packaging information because they anticipate how the discourse will proceed and what content will (or should) be taken up in subsequent utterances. Presupposed information, unlike asserted information, is typically regarded as difficult to address in the ongoing discourse. This is again connected to the projection behaviour of presuppositions. If a speaker utters (15) in response to (11), they are most naturally taken to be denying the assertion of (11) rather than its presupposition. As shown by (13), if we simply negate (11) we allow the presupposition to stand, because it projects from under the scope of negation. Hence, the speaker who responds to (11) with (15) is most naturally understood to mean (13).

(15) That’s not true!

von Fintel (2008), following Shanon (1976), argues that this offers a convenient diagnostic for presupposition: if we wish to deny a presupposition, we have to use a circumlocution such as “Hey, wait a minute…”, as shown in (16) (again considered as a response to (11)). This is dispreferred as a means of addressing asserted content, as shown by (17).

(16) Hey, wait a minute, her boss isn’t away.

(17) ?Hey, wait a minute, she’s not happy.

Given the relative lack of addressability of presupposed content, we might expect cooperative speakers only to presuppose information that they do not expect to be taken up in the following discourse. Otherwise, they would risk giving rise to the sense described by Schwarz (2019, p. 85) that “crucial and important information has been introduced in an inappropriate, underhanded way”. Correspondingly, we might expect a less straightforward and cooperative speaker to be able to sneak controversial information into the discourse without it being questioned, simply by couching that information in terms of presupposition rather than assertion. This assumes that what is paramount for the addressability of the information is its status as presupposed or not – that if material is presupposed, it will automatically be less questionable and addressable than if it had been asserted.

2.3.2 The role of at-issueness

An alternative viewpoint is argued by Simons et al. (2010), who stress the importance of (not-) at-issueness in understanding presupposition projection. On their account, the crucial distinction
is not that between presupposed and asserted content; rather, it is the distinction between material
that is at-issue and that which is not-at-issue relative to the Question Under Discussion (QUD).
The QUD is defined as the question for which the interlocutors are presently committed to finding
the answer Roberts (1996).

As a generalisation, presupposed content tends not to be at-issue, for the obvious reason that
material that is already part of the common ground isn’t usually a good candidate for settling any
open questions. However, in principle, novel presupposed content (for instance, where a speaker
expects to exploit accommodation) can be at-issue, as a speaker could use it to answer the QUD.
Consider the exchange (18a)-(18b).

(18) a. Have you ever worked in Berlin?
    b. I quit my job at the Humboldt University last year.

Taking (18a) at face value as the QUD, (18b) answers indirectly by (formally) presupposing\(^2\) that
the speaker had a job in Berlin. However, this material is clearly at-issue, as it does indeed answer
the QUD, which the non-presupposed content (that the speaker does not currently work at the
Humboldt University) does not. In a similar spirit, there are various politeness formulae that can
be used to introduce novel content but which do so in a way that is formally presuppositional, as
in (19).

(19) Miss Otis regrets she’s unable to lunch today.

Uttered by a waiter to someone sitting in a restaurant awaiting their lunch companion (as in the
Cole Porter song), the main contribution of (19) is to convey that the person in question will not
be attending. Although there is no explicit QUD, the implicit QUD seems more likely to concern
whether Miss Otis will attend than whether Miss Otis regrets anything. Hence, the presupposed
content of (19) appears to be at-issue.

In cases such as (18b) and (19), we could hardly say that the speaker is being ‘inappropriate’ or
‘underhanded’ in the way they introduce new content into the discourse, even though they are
doing so via clearly presuppositional means, from a formal perspective. Yet it is still possible that
using presupposition in this way has consequences for the addressability of the new content in the
subsequent discourse, depending on the extent to which it is at-issueness rather than presupposi-
tionality that determines addressability.

\(^2\)Based on the definition of presuppositions specified in the introduction, ‘formally presupposing’ here and later in
this thesis, refers to content that is triggered by a lexical item, here regret.
We can distinguish two positions on this question that represent the ends of a spectrum of possibilities. If addressability is purely a matter of at-issueness (as the name rather suggests), then whether material was formally asserted or presupposed should be irrelevant to how and whether a subsequent speaker can take it up as a topic of discussion. Note that in these cases asserted content is also present in the discourse turn, and this might still interfere with a subsequent speaker’s attempts to address the presupposed content, potentially requiring them to use a “Hey, wait a minute”-style circumlocution. At the other end of the spectrum, addressability might be purely a matter of the status of the material in terms of whether it is asserted or presupposed, with at-issueness being moot as far as subsequent discourse turns are concerned.

In this paper, we tackle the issue of addressability by presenting an experiment designed to tease apart the contributions of these two factors, at-issueness and presuppositional status. We do so by constructing a scenario in which a (confederate) speaker presents material that is at-issue but which is sometimes couched as assertions and sometimes as presuppositions, and in which the participant is encouraged to identify and rebut the falsehoods in the confederate’s utterances. In this way we will explore, firstly, whether the participant is equally able and inclined to challenge erroneous material when presented as assertion or presupposition (that is, whether the confederate is able to insert controversial material into the discourse by making it presuppositional, controlling for at-issueness), and secondly, whether the status of the challenged material as assertion or presupposition influences its addressability, as measured by the directness with which the participant is able to challenge it, when they choose to do so. Instead of focusing on formulations such as “Hey, wait a minute”, the directness of objections will be estimated by taking into account the number of words uttered. In addition to linguistic effort we will also explore whether objecting to presupposed content is associated with greater cognitive load, building on work by Loy, Rohde and Corley (2018) showing an increase in disfluencies in scenarios involving deception. Cognitive load will be examined by taking into account the number of hesitations and verbal disfluencies identified.

2.4 Experiment [1]

In this experiment, participants role-played a dialogue with a confederate. The scenario was a police interrogation, in which the participant played the role of the detective and the confederate played the role of a suspect in a robbery. Participants were instructed to ask the suspect specific questions and identify and challenge lies in the suspect’s responses. The aim was to investigate whether participants would respond the same way to false information given in the form of presupposition and in the form of assertion, controlling for QUD by ensuring that the same question was asked and the same answer provided in each case.
2.4.1 Materials and design

Participants were provided with instructions which included the cover story and a list of 19 questions which they were instructed to ask in sequence. Eight of these questions were target items in which the confederate’s response contained false content, packaged either in the form of an assertion (four items) or a presupposition (four items), see Appendix A for the full set of items. Participants were randomly allocated to one of two lists of experimental items, which differed only in how the confederate was instructed to respond to these critical items, e.g. the first question was responded to with an asserted falsehood in version 1 and with a presupposed falsehood in version 2, and so on.

The presupposition triggers used represented a wide range of trigger types (stop, know, regret, discover, return, only, to be annoyed, to be happy), reflecting the variability among triggers documented by much prior research (see Schwarz (2019) for recent discussion), which was not a focus of this study. The confederate’s responses to the other 11 filler questions were the same (asserted truths) in both versions of the task. The critical items are included, in both versions, in Appendix A.

Corresponding to each question, the participant had also been provided with a note describing the information currently known to the police, and instructed to challenge any statement that contradicted that information. The confederate’s initial responses were scripted; she was instructed to admit the ‘truth’ if challenged on any point. Participants’ responses were audio-recorded and later transcribed and analysed.

2.4.2 Participants

50 participants (aged 18–39) of which 46% were female were recruited in Edinburgh and paid for their participation. The only criterion was that they should self-identify as native speakers of English.

2.4.3 Results

Across the critical items, participants objected to the false content in 89% of items in which it was asserted and in 79% of items in which it was presupposed. We conducted a mixed-effects logistic regression, postulating a main effect of content type, to examine whether this difference was significant. The model with maximal random effects structure failed to converge and iterative reduction in RE structure yielded a converging model with only by-subject and by-item random intercepts. The model disclosed a significant effect of content type ($\beta = 0.752, SE = 0.297, p = 0.012$ by likelihood ratio test), indicating that false asserted content was objected to more often than false presupposed
content.

For the cases in which participants did object to the content, the length of their response was measured in two ways: by the number of words uttered, and by the number of hesitations or verbal dysfluencies identified. The former measure was designed directly to investigate the claim that presupposed material would be less addressable in the sense of a speaker requiring more words to object to it (as exemplified by the “Hey, wait a minute” test). The latter measure aimed to explore whether there was evidence of greater cognitive load in cases where speakers were obliged to respond to less addressable content, building on work by Loy et al. (2018) showing an increase in dysfluencies in scenarios involving deception.

We conducted two mixed-effects linear regressions, taking as dependent variables the number of words and number of dysfluencies produced, and postulating again a main effect of content type in each case. A model with maximal random effects structure was used to predict the number of words uttered, and a model with by-subject random slopes and intercepts was conducted to predict the number of dysfluencies. There were no significant differences in number of words uttered ($\beta = 0.96, SE = 1.117, p = 0.367$ by likelihood ratio test) or number of hesitations/verbal dysfluencies between conditions ($\beta = -0.037, SE = 0.086, p = 0.66$ by likelihood ratio test), suggesting that no extra linguistic effort was required to object to presupposed content.

2.5 Discussion

Our experiment was designed to investigate whether the presentation of controversial content as presupposition rather than assertion influenced how it was responded to, when controlling for at-issueness with respect to the QUD. The results suggested that, across the board, there was indeed a dispreference for objecting to presupposed content – that is, from a speaker’s perspective, it is possible to forestall objections to false material to a certain extent by making it presuppositional, even in a context in which such objections are socially sanctioned. However, there was little evidence that speakers had difficulty in formulating objections to presupposed content, when they did choose to engage with it: there was no significant difference between responses to presupposed and asserted content with respect to utterance length and disfluencies.

With respect to the first result, we must acknowledge that participants were generally effective in identifying and challenging falsehoods throughout the experiment, and that the majority of false presuppositions did elicit challenges. However, some QUD-addressing false presuppositions were nevertheless allowed to stand, suggesting that presuppositions do tend to be less addressable than
assertions per se.\footnote{There was no specific set of presupposition triggers that particularly allowed their triggered presuppositions to stand.} One possible explanation for this would be that the presuppositional materials are more complex than their purely assertional counterparts, because they contain asserted content that does not transparently address the QUD as well as presuppositional content that does.

One way of testing such an explanation in future work would be to look for systematic differences between participants' behaviour with different presupposition triggers, because triggers vary in the kind of relationship that they encode between the presupposition and assertion, as discussed by Sudo (2012) and Klinedinst (2012). Compare the exchanges (20)-(21) and (22)-(23).

(20) Did Mary argue with her boss?
(21) She regrets doing so.
(22) Did John use to smoke?
(23) He quit recently.

With the trigger regret, as in (21), the presupposition (that Mary argued with her boss) answers the QUD directly, but the assertion (that Mary regrets arguing with her boss) entails the presupposition and hence also answers the QUD. With the trigger quit, as in (23), the presupposition (that John used to smoke) answers the QUD, but what is sometimes taken to be the assertion (that he does not currently smoke) does not answer the QUD.

Consequently, in a regret-type case, one could argue that the presupposed content is not effectively ‘concealed’ as it is also entailed by the assertion, and therefore one would expect a high proportion of challenges to false presuppositions in such a case. In a quit-type case, the presupposed content is independent of the assertion and therefore potentially less salient, and less addressable. However, our experiment does not license us to explore this question in detail as each trigger occurred in just one sentence, risking confounds with item effects.

With respect to the participants' behaviour in cases where they challenge false material, the results appear to support the at-issueness account of Simons et al. (2010). There is no indication that participants felt obliged to use circumlocutions in order to challenge presupposed but at-issue content: these materials, at least in this context, did not appear to elicit “Hey, wait a minute”-style behaviour from the participants. This may be illustrated by taking a closer look at participants' objections towards both false presupposed (24a)-(24e) and false asserted content (25a)-(25e).\footnote{The following dialogue examples begin with the experimental item, i.e. the question-answer pair, see (24a)-(24b) and (25a)-(25b), and are followed by objections of specific participants, named P1, P2, P3 etc., see (24c)-(24e).}
(24) Condition: Presupposed content
   a. Q: Have you held any other positions?
   c. P1: Was that not in Shenzhen China?
   d. P2: That’s not true.
   e. P3: Okay um how long were you in Russia for?

(25) Condition: Asserted content
   a. P: Have you held any other positions?
   b. S: I used to work for the national gallery in Russia until 2017.
   c. P4: Russia or Shenzhen in China?
   d. P5: That’s not true you were working in China.
   e. P6: Why did you leave?

In both conditions, participants object rather directly to the falsehood of the suspect’s claim to have worked in Russia: compare (24c)-(24d) with (25c)-(25d). Hence, the “Hey, wait a minute” test may be mainly sensitive to the informational status rather than the presuppositional status of content. Furthermore, from a qualitative point of view, similar objection strategies were used independently of the content’s presuppositional status: participants objected by asking follow-up questions that addressed the false content (24c)/(25c), by raising the issue that the suspect lied (24d)/(25d), or by asking indirect follow-up questions (25a)/(25e).

Taking both results into account, it seems that in order to predict whether content is available for subsequent discussion warranting discourse coherence one has to account for both the presuppositional status and the at-issueness of content. The approach of Abrusán (2011), further developed in Abrusán (2016), reconciles these two aspects by claiming that although hearers pay attention to certain aspects of meaning by default, their attention may be shifted by contextual cues. Despite being developed for predicting presupposition projection this account seems applicable to this scenario: presupposed content is accommodated by default, but as soon as the hearer’s attention is broadened by contextual cues, the content is available for further discussion to the same extent as asserted content. As regards the potential differences between presupposition triggers, Abrusán (2016) claims that the complements of factives like know can be brought to the focus of attention more easily than the complements of emotive factives like regret, since in the latter case, hearers direct their attention towards the attitude holder instead of the complement’s content. In accordance with the reasoning above, the pre-state implicature of the presupposition triggered by stop is claimed to be focused even less easily, ‘concealing’ the presupposed content more effectively. But
again, differences between presupposition triggers remain to be investigated in future work.

Clearly we should exercise caution about interpreting these results, in that the use of this novel paradigm gives rise to questions about the naturalness of the participants’ elicited behaviour. The kind of objections elicited by the false statements in this paradigm might be atypical for at least two (contradictory) reasons. Firstly, the participants may have been unusually willing to flatly contradict false presuppositions because they were aware that the scenario placed them (playing the police officer) in a position of power relative to the confederate (playing the suspect) and entitled them to change the subject and discuss any issue that they wished to, rather than adhere to the topics foregrounded by the confederate. Secondly, our participants may have been uncomfortable at the task of repeatedly contradicting or challenging their interlocutor and started using circumlocutions and unnecessary politeness formulae when questioning assertions as well as presuppositions. On a similar note, the interrogation setting may have encouraged the participants to actually back up their objections, with concrete evidence which led to longer objections overall, see (26a)-(27c).

(26) Condition: Presupposed content
a. Q: Have you held any other positions?
   c. P7: That’s interesting I have here in my document that you were an employee at the national gallery in Shenzhen in China.

(27) Condition: Asserted content
a. Q: Have you held any other positions?
   b. A: I used to work for the national gallery in Russia until 2017.
   c. P8: That’s interesting cause of right now we have/ at least on my record it says that you were at the national gallery in Shenzhen in China.

Nevertheless, the potential advantage of this paradigm is that it creates a scenario in which repeated false statements are made, each for a clearly-motivated reason, and in which these falsehoods can be challenged naturalistically without violating politeness norms.

2.6 Conclusion

The experimental results presented in this paper suggest that, when material is at-issue, it can be challenged directly by a subsequent speaker whether it is formally asserted or presupposed. However, expressing at-issue material through presupposition rather than assertion appears to have the effect of reducing the frequency of such challenges. These findings are consistent with a view
on which speakers are able to manipulate their interlocutors’ ability to address discourse content to some extent through the formal apparatus of presupposition, but where material that is relevant to the Question Under Discussion is usually available for subsequent challenge to quite a pronounced extent. Thus, a speaker-hearer model that predicts what material is eligible to discuss in the subsequent dialogue must account both for interlocutors’ expectations about information packaging as well as about the overall discourse topic.

2.7 Bayesian Analysis

2.7.1 Objection rate

Since the remaining experiments of this thesis are analysed in the Bayesian framework, we will analyse the previously discussed data again in order to keep the analysis coherent. Bayesian models were chosen throughout this thesis because they can fall back to prior information which enhances convergence. Frequentist models with maximal random effects structure do often not converge as was the case for experiment [1], see section 2.4.3 above. This led us to use a reduced effects structure which might have caused an inflated false positive effect of content type on the probability to object (Barr, Levy, Scheepers & Tily, 2013). Thus, re-analysing our data in the Bayesian sense will also explore whether the found effect of content type is indeed reliable. The data was analysed 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).

To analyse the objection rate, a Bayesian binary logistic regression model with maximal random effects structure was conducted. The experimental factor content type (Assertion/Presupposition) was included to predict the probability to object. The factor content type was sum-coded: 1 as Assertion and -1 as Presupposition. The model included varying intercepts and slopes for participants and presupposition triggers, assuming that the effect of content type on the participants’ objection rate varies between participants and presupposition triggers.

Weakly regularising priors were used, which allowed a reasonably wide range of parameter values and at the same time penalised very extreme values. The prior for the intercept was a normal distributions with mean 0 and standard deviation 10. This means that one could be 68% certain that the intercept would fall within -10 and 10 on the log-odds scale which translates approximately to a range of 0 and 1 on the probability scale. For the fixed effect of content type, a normal prior with a mean of 0 and a standard deviation of 1 was 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).

Samples were drawn from the posterior distributions of the model parameters using the NUTS sampler (Hoffman & Gelman, 2013). Four sampling chains were run, each collecting 4000 iterations whereby the first 1000 iterations were disregarded as part of the warm-up phase leading to 12000 iterations available for analysis.

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. One could roughly interpret the evidence as reliable if zero lies outside the parameters’ 95% credible interval (Kruschke, Aguinis & Joo, 2012).

Results

The model disclosed a weak effect of content type ($\hat{\beta} = 0.34$, CrI: [-0.12, 0.77]), indicating that false asserted content was objected to more often than false presupposed content. However, this effect is not reliable: zero lies within the parameters’ 95% credible interval, illustrated in Figure (2.1). To be more precise, the probability that the true parameter value is greater than zero is 93%. Recall that when we analysed the same data within the frequentist framework using a model with reduced random effects structure, we found a significant effect of content type. The Bayesian analysis reveals that this effect is weak and must be interpreted with caution.

2.7.2 Circumlocutions

Corresponding to the frequentist analysis, we investigated (i) whether participants required more words to object to presupposed than to asserted content by looking at the length of the participants’ responses (total number of words without filler words); and (ii) whether there was evidence of greater cognitive load when participants responded to presupposed content opposed to asserted content (number of hesitations and/or disfluencies).

For this purpose two mixed-effects Bayesian models were conducted, both postulating a main effect of content type. In order to predict the number of words, a mixed-effects negative binomial regression model was carried out with a log link (linguistic effort model). The negative binomial regression model better accounted for the over-dispersed outcome variable number of words, while
the corresponding Poisson model underestimated the dispersion of the data.

Weakly regularising priors were used. The prior for the intercept was a normal distribution with mean 2.3 and standard deviation of 1.6. This approximately translates back to a mean of 10 and a standard deviation of 5 on the original scale and is based on Nippold and Hayward-Mayhew (2013)’s interview data which showed that participants used on average 10 words per utterance, ranging roughly from 6 to 16 words. Thus, given our prior one could be 68% certain that the intercept would fall within 5 and 15 words. For the fixed effect of content type, a normal prior with a mean of -3 and a standard deviation of 1.6 was used. This approximately translates to a mean of 0 and a standard deviation of 5 on the original scale. Thus, one could be 68% certain that the effect of content type – i.e. the difference between objecting to asserted content and presupposed content – is less than 10 words.

A mixed-effects binomial regression model was used to predict the number of hesitations/disfluencies within a participants’ response (cognitive effort model). Weakly regularising priors were used, which allowed a reasonably wide range of parameter values and at the same time penalised very extreme values. The prior for the intercept was a normal distribution with mean 0 and standard deviation 10. For the fixed effect of content type, a normal prior with a mean of 0 and a standard deviation
of 1 was used.

For both models, 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. The same sampling process as before was used, see 2.7.1. Both models included varying intercepts and slopes for participants and presupposition triggers, assuming that the effect of content type varies between participants and presupposition triggers. The predictor content type was sum-coded: 1 as Assertion and -1 as Presupposition.

**Results**

Focusing first on the linguistic effort, objections were on average 13.5 words long, with a standard deviation of 10 ranging from 1 to 68 words which is close to Nippold and Hayward-Mayhew (2013)'s findings. The linguistic effort model did not disclose any effect ($\hat{\beta} = -0.05$, CrI: [-0.14, 0.03]). Turning to the cognitive effort in terms of verbal disfluencies, objections contained on average 1.5 disfluencies with a standard deviation of 1.91 ranging from 0 to 8 disfluencies. Recall that we counted filler words, hesitations, repetitions and repairs as disfluencies. The cognitive effort model did not disclose an effect of content type ($\hat{\beta} = -0.02$, CrI: [-0.31, 0.27]).

**2.8 What’s next?**

The Bayesian analysis could partly verify the results of the frequentist analysis: We found a weak effect of content type on the probability to object to content, such that speakers were more likely to object to false asserted content than false presupposed content. However, this was only a weak effect, and thus, has to be investigated further. Furthermore, we did not find any indication that objecting to presupposed content required more linguistic or cognitive effort than objecting to presupposed content. As was stated earlier, not having found any indication that circumlocution were necessary to object to presupposed content suggests that the “Hey, wait a minute” test may be mainly sensitive to the informational status rather than the presuppositional status of content. Crucially, these results are compatible with Potts (2008)'s findings on the usage of the actual discourse move “(Hey) wait a minute” in a corpus of CNN TV interview transcripts. His findings show that “(Hey), wait a minute” is not only used to object to presupposed content but a wide variety

---

5 A second analysis was carried out using less informative priors. The model with less informative priors yielded the same result.

6 A replication experiment for the first experiment was planned, however, the pandemic made in-person testing impossible.
of content that is not at-issue. Thus, the broad “Hey, wait a minute” test seems insufficient in determining whether content is presupposed or not.

The remainder of this thesis will investigate the distinction between presupposed and asserted content further by focusing on specific presupposition triggers, more specifically, the factive presupposition trigger *know* (Part II) and the additive presupposition trigger *too* (Part III). The centre of the investigation will be the production of these triggers, which, to our knowledge, has not been investigated yet. We decided on factive presupposition triggers for Part II partly based on the data of experiment [1]: *know* and *discover* showed the observed (but weak) pattern of higher objection rates when content was asserted (0.84 and 0.83 respectively) opposed to when it was presupposed (0.77 and 0.80 respectively). Furthermore, factives constituted a practical starting point for testing a forced-choice paradigm which requires presenting participants with a set of alternative formulations of which *know* has plenty, notably including: (i) verbs that take sentential complements and convey the speakers’ degrees of belief, e.g. *believe*, *think*, *guess* etc., and (ii) the bare assertion.

Part II will first compare the usage of *know* versus the aforementioned sentential verbs which will be call (un)certainty expressions, and second contrast *know* with the bare assertion. Thus, Part II will shed more light on *know* as an expression that conveys degrees of belief [experiments 2 and 3], as well as its role as presupposition trigger and potential discourse structuring tool that may be used to emphasise what content is being picked up in the preceding dialogue [experiments 4 and 5].

---

7Note, that I only compare relative frequencies and do not discuss the results of any statistical tests here. The differences between presupposition triggers was not investigated statistically, since the effect of trigger type may be confounded with items effects.
Part II

Discourse Structuring
Chapter 3

The difference between knowing and believing

3.1 Abstract

Speakers have a number of options when introducing propositions which they take to be uncertain: for instance, they can use verbs such as know, believe or think. The production of uncertainty expressions is highly context dependent. One promising approach to capturing the semantic meaning of these expressions takes them to be available only when the speaker’s confidence in the proposition exceeds some threshold. However, it is unclear whether this approach deals satisfactorily with the full range of usages of uncertainty expressions. For instance, speakers may also use them to achieve social goals such as toning down the force of their assertion. In this case they pursue another communicative goal than just being cooperative: they also aim to be polite. The current study investigates the speakers’ motivations in choosing between uncertainty expressions such as believe or the factive know in two controlled contexts. More specifically, we show that speakers’ choice of expression is influenced by (i) how likely they estimate an event to be and (ii) strategic considerations relating to the communicative context in which they are working. Thus, speakers adjust their language as a manipulative process. We situate these results in the context of threshold semantics.

3.2 Introduction

Speakers have access to a vast repertoire of tools, such as uncertainty expressions (e.g. believe, think) and factive verbs (e.g. know, notice), to convey their degrees of belief about a specific state of the world, or whether a particular event has taken place. Uncertainty in communication and
interaction has been theorised from different perspectives (e.g., Littlejohn, Foss & Oetzel, 2017), several of which address the idea that the communicator’s goal in interaction is often to reduce or manage their (cognitive) uncertainty (e.g., Berger & Calabrese, 1975; Berger, 1995; Brashers, 2006). Work in linguistic semantics and pragmatics has paid particular attention to the use of expressions that convey information about the (un)certainty of propositional information, which constitute an important tool for reducing a hearer’s uncertainty as to the current state of affairs in the world. Among these expressions are verbs which take sentential complements and which convey different degrees of speaker confidence in the factuality of those complements.

Factive verbs such as know are argued to presuppose the truth of their complements, under which assumption we might expect them to be used only by speakers who are certain about the factuality of those complements. By contrast, verbs such as believe convey no such presupposition. For example, a speaker uttering (28a) might only have plausible reason for thinking that the glasses are on the kitchen table. In contrast, in (28b) the speaker seems to convey certain knowledge about the location of the glasses. Focusing on the contrast between (28a) and the bare assertion in (28c), the uncertainty expression in (28a) can be understood to convey the speaker’s confidence (or the lack thereof) in the truth of the proposition that the glasses are on the kitchen table.

(28)  
a. I believe that you left your glasses on the kitchen table.
   b. I know that you left your glasses on the kitchen table.
   c. You left your glasses on the kitchen table.

Recent work attempting to formalise the semantics of uncertainty expressions (e.g., Yalcin, 2010; Lassiter, 2017) has appealed to the idea of thresholds of probability: For any expression there exists some threshold in the range \([0, 1)\), and an utterance containing that expression is true if the probability of the event it describes exceeds this threshold. Given that cooperative speakers are expected to provide as much relevant information as they can (following Grice, 1975), we would pragmatically expect a speaker to choose the utterance with the highest threshold they can: that is, they should utter (28b) rather than (28a) if they hold that the probability of the event exceeds the threshold that would make (28b) true.\(^1\) Hence, the hearer of (28a) might infer that the speaker is not certain enough to assert (28b).

However, this inference relies on the assumption that there is no reason for a speaker to have used a weaker alternative: that is, to have said believe when they were in a position to say know. In

---

\(^1\)Note that this kind of reasoning only applies to those (un)certainty expressions that impose a lower bound (such as the verbal (un)certainty expressions discussed here) but not to (un)certainty expressions that impose an upper bound (e.g. uncertain).
practice, we might intuit that a speaker will sometimes opt for the weaker option despite their knowledge state. One possible motivation for this would be politeness. For example, suppose that a valuer has inspected a painting which the owner thinks is worth a fortune, and the valuer is certain that it is not. Out of politeness, they might still utter (29a) rather than (29b).²

(29) a. I believe that your painting isn’t worth much.
   b. I know that your painting isn’t worth much.

This paper will examine language users’ expectations about the choice between know and uncertainty expressions such as believe. In particular, we investigate the effect of the strength of evidence available to the speaker, across two different scenarios in which the expectations of appropriate speaker behaviour differ. As such we examine ways in which participants estimate speakers’ strategic use of (un)certainty expressions, or to phrase it differently, ways in which they posit that speakers may be using language as a manipulative tool. The broad aim is to explore the usefulness of a threshold semantics approach in capturing language usage in this domain. Experiment 1 [2] focuses looks at a variety of verbs conveying different degrees of belief: know, notice, to be sure, think, believe, see, guess. Henceforth, the term (un)certainty expressions will be used when referring both to uncertainty expressions and to factive verbs. Experiment 2 [3] will zoom in onto the (un)certainty expressions believe and know.

3.3 Previous work on (un)certainty expressions

3.3.1 Threshold semantics

Uncertainty expressions indicate the degree to which speakers are committed to the content of their utterances and reveal the speakers’ knowledge about the truth of the presented proposition. Interlocutors rely on contextual factors to produce and interpret uncertainty expressions, since there is no straightforward translation between uncertainty expressions and event probabilities (Clark, 1990). For example, hearers have been found to have difficulties in understanding verbal probability descriptors such as ‘common’ when being informed about the risk of medical side effects (Knapp, Raynor & Berry, 2004).

One way of formalising uncertainty expressions, semantically, would be to assume a threshold semantics (Lassiter, 2017; Yalcin, 2010): For any expression there exists some threshold [0, 1), and

²It has to be added that when speakers convey subjective/unknowable material (e.g. their stance on whether god exists), (un)certainty expressions associated with relatively low degrees of belief such as believe or think seem to be the only warranted options (e.g. ‘I believe/think that god exists.’) even if the speaker is (subjectively) extremely certain.
an utterance containing that expression is true if the probability of the event it describes exceeds this threshold. For example, if the threshold a given speaker in a specific situation has for believe is 0.6, then the utterance ‘I believe that Scotland will win the match tomorrow’ is available for the speaker if they believe that the probability of Scotland winning the match tomorrow exceeds the threshold of 0.6.

However, there is no straightforward translation between a given formulation and the probability that the event described by the propositional content is true (Clark, 1990), speakers’ production choices and hearers’ interpretations of (un)certainty formulations rely on contextual factors. This kind of context-sensitivity of (un)certainty formulations can be accounted for within thresholds semantics by assuming that thresholds are distributional rather than fixed values. Possible contextual information that interlocutors take into account when producing and interpreting uncertainty formulations have been found to be the distribution of alternative outcomes. For example, in a study by Lassiter (2017), participants were asked to agree or disagree with the sentences x might win the raffle. in different scenarios: e.g. in scenario (1) 1000 individuals had one ticket each, in scenario (2) six individuals had one ticket each, while one individual had 994 tickets. Despite that x was always depicting an individual holding one ticket with the same probability to win, participants were found to agree with the statement more readily in scenario (1) than in (2). These result suggest that the associated threshold for might was on average lower in scenario (1) than in scenario (2). This was similarly true for possible in It is possible that x will win the raffle., whereby the threshold of possible was generally lower than the one of might.

Furthermore, the usage of (un)certainty expressions was found to have considerable amount of interspeaker variability (e.g., Wallsten, Budescu, Rapoport, Zwick & Forsyth, 1986). In a recent study, Schuster and Degen (2020) suggest that hearers use the speaker’s identity when interpreting utterances: what expressions does a speaker preferably use and what thresholds do they associate with specific expressions? Schuster and Degen (2020)’s study tests how hearers adapt to speaker-specific thresholds when interpreting uncertainty expressions. In one part of their study, participants were asked to listen to either a cautious or confident speaker producing a bare assertion or utterances containing probably or might. The confident speaker would use might and probably to describe lower event probabilities than a cautious speaker. Participants were then asked to make guesses about the production of the speaker they were introduced to in the first part choosing between might, probably and a something else option.

The results show that participants in the confident speaker condition gave high ratings for probably for a larger range of event probabilities than for probably in the cautious speaker condition. For
*might* the opposite was observed: participants gave high ratings for *might* for a larger range of event probabilities in the cautious speaker condition than in the confident speaker condition. In a third experiment, participants inferred higher event probabilities for *might* and *probably* when produced by the cautious speaker than the confident speaker. Thus, hearers seem to be able to adjust to speaker-specific thresholds, e.g. use higher thresholds for expressions such as *probably* when interpreting utterances communicated by a cautious speaker.

Schuster and Degen (2020)’s findings suggest that hearers adapt to the speakers’ identity. Crucially, this line of research focused on the utility of an utterance as being determined primarily by its informativeness and the speakers’ utterance preference. Thus, speakers aim to reduce uncertainty for the hearer by choosing the most informative utterance while also having personal preferences as to which utterance to choose. As was pointed out by Schuster and Degen (2020), while these considerations successfully capture inter-speaker variability, it is not clear how such an approach would capture the effect of additional communicative goals speakers may have when using uncertainty expressions. For example, a speaker may use *might* to make a statement more moderate, or to be polite – goals which may vary by context even for the same speaker and which conflict with the goal of presenting the strongest possible information, in order to reduce uncertainty for the hearer. In the following section, these goals will be discussed in more detail.

### 3.3.2 Communicative strategies

**Politeness**

It has long been noted that uncertainty expressions that convey modal meaning also give rise to particular discourse effects (e.g., Fraser, 1975; Holmes, 1982). Uncertainty expressions may function as *downtoners* (Holmes, 1982) or *hedges* weakening the assertive strength of an utterance, and yielding discourse effects such as vagueness or politeness. The concept of *hedges* was popularised by Lakoff (1973) to capture linguistic expressions, such as *sort of*, that can signal different degrees of category membership of a particular expression, see (30).

(30) This paper is sort of long.

This type of hedging was later called *propositional hedging* (Fraser, 1975), since the hedge affects the truth value of the propositional content. Hedging was further investigated by Brown and Levinson (1978) who extended hedging to expressions that modify the speaker’s commitment to a proposition, such as *think* (sometimes discussed as *speech act hedges* (Fraser, 2010)). This type of hedging is the focus of this paper. For example in (31), a teacher might want to gently introduce the subject about a pupil having to repeat the 4th grade to the parents (31a) than to directly assert it (31b).
(31)  a. I believe that your son will have to repeat the 4th grade.
    b. Your son will have to repeat the 4th grade.

Within politeness theory (Brown & Levinson, 1987), being polite is often analysed through the lens of facework: that is, the interlocutors’ aim to maintain their positive or negative face. 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 teachers are discussing a student’s mark, see (32), Teacher2 is threatening their colleagues’ positive face by criticizing them. Similarly, this situation could be potentially face threatening for Teacher2 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 (32a) – i.e., hedging – instead of asserting the criticism directly (32b).

(32) Teacher1: This essay was submitted on time.
    a. Teacher2: I believe you are mistaken; the student submitted the essay a day late.
    b. Teacher2: You are mistaken; the student submitted the essay a day late.

Empirical research suggests that speakers indeed consider motives such as politeness when communicating their degrees of belief. Juanchich and Sirota (2015) investigated the way in which speakers communicate more or less face-threatening news to their friend. More specifically, participants were confronted with a scenario that had less severe consequences (friend’s car breaking down), and another that had more severe consequences (friend making a bad investment). Both events were characterised as being 50% likely to occur. Juanchich and Sirota (2015)’s findings suggest that tactful speakers who are concerned about the hearer hedge their utterances, understating their confidence by using expressions such as a very small probability, a small probability, slightly probable.

Holtgraves and Perdew (2016) extended this line of research by looking at scenarios with varying event probabilities (20%, 50% and 80%) from the production and comprehension perspective. The findings of a production task indicate that speakers hedge their utterances by using expressions conveying lower degrees of belief when hearers are more severely affected by the event the speakers are describing (e.g. ‘It’s somewhat unlikely/likely that the car needs a new transmission.’). This was the case even when both events – severe and less severe – were equally likely.3

From the hearers’ perspective, Holtgraves and Perdew (2016) found that participants assigned high probabilities (value ranging from 0 to 100) to a given event when hearing expressions conveying high

---

3 Another way of explaining these results would be to assume that severe events have lower prior probability than non-severe events which speakers may take into account in their production choices.
certainties, such as ‘definitely’ (e.g. ‘The car definitely needs a new battery/transmission.’). At the same time, participants assigned lower probabilities to severe events (car needs a new transmission) than for less severe events (car needs a new battery). This is a striking result in the context of prior research by Bonnefon and Villejoubert (2006) showing 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 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). At the same time, speakers are also perceived as warmer, as we might predict if hearers take into account politeness theoretic constraints.

**Power dynamics**

The severity of a face threatening action depends partly on the power relation between the interlocutors (Brown & Levinson, 1987). Low-power speakers may be more concerned about facework than speakers who have the same or a higher social status than the hearer. For example, if a teacher were talking to their student, as in (33), rather than a colleague, as in (32), they might be less inclined to resort to hedging (33b) and more inclined to contradict the speaker directly (33a).

(33) Student: I submitted my essay on time.
   a. Teacher: You are mistaken; you submitted the essay a day late.
   b. Teacher: I believe you are mistaken; you submitted the essay a day late.

The previously discussed studies show that speakers use hedging when they are on a par with the hearer (communicating with friends, Juanchich and Sirota (2015)) or in a lower social position (communicating with parents, Holtgraves and Perdew (2016)). It is less clear to what extent high-power speakers engage in facework. From the hearers’ perspective, high-power hearers (role of parent) may dismiss the possibility that the low-power speaker (child) could use hedging (Holtgraves & Perdew, 2016). On the other hand, as shown by Bonnefon and Villejoubert (2006) in the context of medical communication, low-power hearers (patients) appear to be aware that a high-power speaker (doctor) may hedge their statements.

However, the effects of power relations might arise here for two separate reasons: because the bare assertion is taken to be face-threatening to the hearer in a way that is inappropriate in the presence of a particular power dynamic, or because the risk to the speaker of making a false statement is higher in such a case. That is, it’s not obvious whether the speaker who hedges does so because they could confidently make a stronger statement but that would be inappropriate, or because
they require higher confidence to make a stronger statement in the presence of an unfavourable power dynamic. In a threshold semantics analysis, this latter case could be treated as a case in which the threshold for making the stronger statement has increased. For example, immediately after the vote count for the 2020 US presidential election, Republican senators who refused to say that Trump had lost may have done so because they feared reprisals for saying so even though they thought it was certainly true, or because they did not want to take the risk of asserting his defeat falsely in a circumstance where they thought he might still have a very slim chance of winning.

The previous point may also be looked at from the perspective of commitment – whereby the commitment to a proposition $p$ is understood as ruling out the alternative $\neg p$ (Gunlogson, 2004). Following Gunlogson (2004), speakers can commit to a proposition and thereby make it a public belief. Such a belief is not necessarily mutually shared between discourse participants which is the case when all discourse participants jointly commit to a proposition. Committing to a given proposition when the belief may not be mutual or even committing to a controversial proposition, i.e. when other discourse participants share the public belief that $\neg p$, may require the speaker to be in a favourable power position. Furthermore, in order to make the belief a mutual one, the other discourse participants have to ratify the proposition. Without such a ratification, the proposition remains only public but not mutual. One may argue that the addressee has to be in the adequate power position to ratify such a proposition. Back to the example of the Republican senators, one may argue that the Republican senators were not in the right power position to commit to Trump having lost the election or did not have the evidential authority to ratify such a claim.

Previous studies have singled out specific power dynamics. One way of interpreting these findings is as evidence of speakers’ selection of communicative strategies. Criticising someone who is in a high-power position requires a different strategy than criticising someone who is in a low-power position. A successful strategy for a low-power speaker could be to hedge/downplay their certainty about their interlocutor’s mistake. However, there are scenarios where a speaker might choose a quite different strategy: conveying high certainties. For example, in a scenario where the hearer is thought to have lied about what they did last night and the speaker has at least a suspicion about what the hearer was up to, they might want to act as if they are already certain about what the hearer did. By doing so the hearer may be more inclined to tell the truth, assuming that the hearer believes the speaker’s certainty to be as high as suggested by the speaker’s utterance and does not anticipate a bluff. Ways of conveying high certainty would include using the bare assertion, see (34b), or to use know (34a).

(34)   a. I know that you went to the party without me.
b. You went to the party without me.

Note that the use of the factive presupposition trigger *know*, although indirect, might confer an advantage over using the bare assertion: by using *know* the speaker not only conveys that they are highly certain that the hearer went to the party without them but also acts as if this content is a fact and something that everyone including the hearer can agree on. By definition, presuppositions are presumably already shared knowledge. If presupposed content happens to be new, one way of repair is for the interlocutors to accommodate the presupposition: that is, to act as if it was in the common ground. Presupposed content triggered by *know* – the interlocutor being at the party without the speaker – might be less likely to be challenged by the hearer, even if it is in fact controversial. *Lorson, Cummins and Rohde (2019)* (Chapter 2) contrasted the usage of assertions versus presuppositions and the findings suggested that (formally) presupposed content was less likely to be challenged by the hearer than asserted content.

In summary, speakers use uncertainty expressions to communicate their degrees of belief. The willingness to use particular expressions seems to vary between speakers, and interlocutors are flexible in adapting to each other’s production preferences. But, at the same time, speakers may also use a certain expression when the event they are describing does not exceed the threshold of that expression. The thresholds of uncertainty expressions seem to vary between interlocutors and interlocutors seem to be flexible when it comes to adjusting to each other’s way of using uncertainty expressions. On top of that, speakers may use uncertainty expressions as hedging devices or downtoners following communicative goals such as being polite. Thus, it seems that there are a multitude of factors contributing in different ways to the speakers’ production of uncertainty expressions. In this paper, we will explore whether we can appeal to a threshold-based semantic analysis to systematise the effects of these competing considerations.

### 3.4 The present study

The current study investigates the speakers’ motivations in choosing between uncertainty expressions such as *believe* or factive verbs such as *know*. More specifically, we explore whether participants’ choices of expressions are influenced by (i) how likely they estimate an event to be and (ii) strategic considerations relating to the communicative context in which they are working. We will extend previous research by introducing a different way of assessing participants’ degrees of belief, and by introducing a within-subjects manipulation that examines the effect of context on a speaker’s strategic utterance choice.
The majority of studies that were discussed have used quantitative prompts to manipulate the probability that an event takes place. Using probabilities is problematic for three reasons: (i) In daily life speakers don’t usually know the exact probability with which an event takes place; (ii) speakers rely on evidence/arguments/experience/intuition rather than reasoning about event probabilities when communicating degrees of belief; (iii) interlocutors usually perform quite poorly when it comes to understanding probabilities (Kahneman, 2011). In this study, participants are asked to choose between utterances to produce, while showing them pictures and documents that can be implicitly evaluated for the event certainty they denote. The participants’ degree of belief were assessed after the production task by asking participants to rate how likely they thought it was that a specific event took place given a piece of evidence. By taking this approach, we aim to elicit production choices and event certainty judgements that are more similar to those occurring in daily communication.

Prior work has shown that participants can adjust to speaker-specific utterance thresholds (Schuster & Degen, 2020). Here the question is whether speakers’ use of (un)certainty expressions also varies depending on the scenario they are in and who they are talking to. We target two scenarios that vary in their power dynamics to compare production choices made by a speaker in a high-power position to those made by a speaker on a par with their interlocutor. The advantage of a within-subjects manipulation (in contrast to prior work that has tested only one power dynamic or another) is that we can examine the extent of context-specific adaptation while holding the speaker constant. In this study, we tested whether participants adapt to a change of context to change the way they convey their (un)certainty.

Two experiments were conducted – the first experiment examines seven (un)certainty expressions and the second experiment focuses on know versus believe. Both involved a production task where participants were asked to choose between utterances to convey messages about different events, followed by an evaluation task where participants were asked to adjust a slider to indicate their evaluation of the certainty of those events. Note that, while we characterise this primarily as a production task, participants’ choices in this task can also be understood to involve elements of comprehension: the participant is asked to indicate what a character in a scene is likely to say, by evaluating and choosing among different candidate utterances. This decision-making process

---

4The ability to deal with magnitudes and numerical operations is part of what Kahneman (2011) refers to as system 2. System 2 incorporates all the cognitive processes that require effort and time. So for one part counting, estimating and understanding numerical information is effortful. In contrast, judging pieces of evidence is based on intuitive and fast reasoning (system 1). Furthermore, independent of Kahneman (2011), innumeracy (the struggle to deal with magnitudes and numerical operations including understanding percentages/probabilities) is a widespread problem that comes with societal costs (Foundation, 2014, e.g.). A YouGov poll showed that out of 2,000 UK citizens, only 26% could solve a simple percentage problem about the sugar content of a chocolate bar (Foundation, 2016).
reflects language users’ awareness that what they say is guided by how it may be understood by the hearer, meaning that speakers and hearers engage in what can be termed ‘mutual vigilance’ (Sperber et al., 2010). The only difference between these experiments in terms of task is the number of utterances participants had to choose from. The full set of data for both experiments is available here: https://osf.io/e5av9/.

3.4.1 Experiment 1 [2]: seven (un)certainty expressions

The first experiment investigated the production of a wide range of (un)certainty expressions (know, notice, to be sure, think, believe, see, guess) in two controlled, contrasting scenarios. The goal of this experiment was to (i) test the experimental design and the materials and (ii) narrow down which (un)certainty expressions we should contrast in the second experiment.

The experiment consisted of two tasks. For the production task, participants were asked to play the role of a detective in an investigation of an art heist where they briefed a colleague and interrogated a suspect, relying on pieces of evidence about the suspect’s whereabouts. In this way we elicited expressions that correspond to degrees of belief without providing explicitly quantitative prompts. After the production task, participants then evaluated their confidence in each piece of evidence retrospectively. This evaluation is referred to here as the evidentiality measure. Thus, participants produce utterances in two scenarios corresponding to a range of confidence levels in the propositions uttered.

We expected participants to be pragmatically cooperative in the sense of producing the most informative statement that they truthfully can. If the threshold-based account of the semantics of these verbs is correct, we would expect speakers to select the verb with the highest threshold that does not exceed their degree of belief in the complement proposition. This means that, assuming know has a higher threshold than believe, speakers are predicted to choose believe over know as long as the probability of the event they are describing does not exceed the threshold for know. Following Schuster and Degen (2020)’s results, we expect speakers to differ in their thresholds for producing specific expressions.

Considering work in the domain of politeness theory, we generally predict that verbs expressing higher degrees of belief will be more widely used in the interrogation than in the briefing. We examine whether this difference can be captured by assuming that the scenario exerts a general effect on the thresholds for producing particular utterances. This is based on the assumption that speakers might have a systematic tendency to downplay their certainty when speaking to their colleague in a cooperative scenario, compared to when they are interacting with a suspect.
in an uncooperative scenario. This could also be reformulated in terms of power relations: In the interrogation the speaker is in a high-power position compared to the briefing, which potentially obviates their need to consider politeness constraints.

Participants

35 participants were tested, recruited over the crowd-sourcing platform Prolific, specifying participants with an approval rate above 90. Participants were paid an average of £7.53/hour (the average duration of the experiment was 35 minutes). The age of the participants ranged from 18 to 52 years, with a mean of 24 years. 23 participants stated their preferred pronoun as she/her, 11 chose he/him and 1 chose they/their.

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 estimated to be weak in evidentiality or evidence that we considered to be strong in evidentiality. However, for the analysis, we did not rely on this categorisation of evidential strength, but instead on the participants’ evaluation of that in a post-test (see below). In order to ensure that participants saw a particular item only once but at the same time were exposed to both scenarios, we introduced two suspects. In this way, participants interrogated suspect1 and briefed about suspect2 or the other way round. This yielded 40 critical 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 Appendix B for full details. The question in the heading of the picture was meant to be an already asked question by the participant. The answer to the question was provided by either the colleague or suspect, depending on the scenario, in form of a speech bubble. The participants were asked to react to the colleague’s/suspect’s answer by filling the gap in a sentence by choosing between know, notice, to be sure, think, believe, see, or
guess, highlighted in the briefing (35) and (36) items. Alternatively, they were able to choose the option ‘other’ and formulate their own utterance. The order in which the expressions were displayed was randomised.

(35) Briefing item:
Participant: Did Emily Brown have any financial problems?
Colleague: Financially the suspect was doing alright.
Participant: I _______ that the suspect was in need of money.
[know|believe|notice|amsure|think|see|guess]

(36) Interrogation item:
Participant: Did you have any financial problems?
Suspect: Financially I was doing alright.
Participant: I _______ that you were in need of money.
[know|believe|notice|amsure|think|see|guess]

The manipulated pieces of evidence ranged from pictures to 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 (35) and (36) a bank statement was provided as strong evidence and a statement of a friend who mentions potential, financial difficulties as weak evidence; see Appendix B 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 and the participant had to choose the correct answer (here 11am), see (56). The option ‘other’ was also available.

(37) Control item:
Picture: Clock in rooms says it is 11am.
Participants: Oh look at the time. Is it already _______?
[11am|noon|2pm]

After the production task, participants 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. 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. Each piece of evidence dealt with a separate proposition which is why we assume that evidentiality
ratings will not be influenced by anything else taking place in the experiment before participants were asked to rate the evidence. To communicate their certainty, participants adjusted a slider from 0 (not at all certain) to 100 (very certain).

**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. We also informed them about the structure of the experiment: (1) production task, engaging in two discussions, (2) rating evidence, (3) demographic questionnaire. Then the task was introduced in the form of a story about an art heist in Edinburgh involving two suspects. Since the lead detective on the case went missing, the participants were asked to help out solving the case. Both scenarios were introduced by stating that one of the suspects had been arrested. The 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). Participants were instructed to converse with the colleague/suspect about different topics including questions about the case and they were 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 then 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. The evaluation task was followed by the voluntary, demographic questionnaire. The experiment lasted approximately 35 minutes.

**Analysis**

The data was analysed fitting a Bayesian categorical regression model with maximal random effects structure 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 Bayesian framework was chosen because the models with maximal random effects structure did not converge in the frequentist framework. Since Bayesian models can fall back to prior information they converge more easily. We chose not to analyse the data with a reduced effects structure within the frequentist framework to prevent inflated false positive effects (Barr et al., 2013).

The experimental factor scenario (Briefing/Interrogation) and the continuous variable evidentiality ([0, 100]) were included to predict the probability of choosing believe/notice/am sure/think/see over know. This makes know the reference category of the model. The factor scenario was sum-coded: -1 as Interrogation and 1 as 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’ utterance choices varies between participants and items.

Weakly regularising priors were used, which allowed a reasonably wide range of parameter values and at the same time penalised very extreme values. The priors for the by-expression intercepts were normal distributions with mean 0 and standard deviation 10. This means that we could be 68% certain that the by-expression intercepts would fall within -10 and 10 on the log-odds scale which translates approximately to a range of 0 and 1 on the probability 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 et al., 2016).

Samples were drawn from the posterior distributions of the model parameters using the NUTS sampler (Hoffman & Gelman, 2013). Four sampling chains were run, each collecting 4000 iterations whereby the first 1000 iterations were disregarded as part of the warm-up phase leading to 12000 iterations available for analysis.

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 could roughly interpret the evidence as reliable if zero lies outside the parameters’ 95% credible interval (Kruschke et al., 2012).

The response ‘other’ was excluded from the analysis. We included ‘other’ mainly to give participants more freedom in their utterance choice and to create a more natural experience. The response ‘other’ made up only 8% of the data.

Results

The accuracy of participants was at 95% for the control items which suggest that they paid attention during the experiment. Overall, participants used a wide range of expressions, see Table 3.1 for details. guess was attested much less frequently than the other options, so it was omitted from the detailed statistical analysis.
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 71.5. The means, standard deviations and medians of the evidentiality measures for the scenarios briefing and interrogation were fairly similar (mean=72/70, sd=23.9/25.12, median=75/74 respectively), suggesting that the pieces of evidence in the briefing were similarly perceived by participants as the pieces of evidence in the interrogation. This was desired, since otherwise the scenario manipulation would have been confounded with differences of evidence pieces between scenarios. If participants had assigned much higher evidentiality ratings to evidence in the briefing than in the interrogation, the result would have agreed with the intuition that there are higher standards of evidence in the briefing than in the interrogation. Thus, we would not have been able to exclude whether participants used the (un)certainty expressions differently between scenarios because they followed different communicative goals or whether the differences in mean evidentiality ratings for each expression mirrors that a higher threshold was needed in the briefing than in the interrogation. However, considering the way we tested evidentiality, it makes sense that the evidentiality ratings for each scenarios are almost identical, since participants were asked to rate the pieces of evidence for both scenarios together in a randomised order after the production task. Only the participants’ memory would have made it possible to associate a given piece of evidence with either the briefing or the interrogation.5 With regards to differences between formulations, participants used know when they saw fairly convincing evidence (mean evidentiality rating was 81), whereas they used think for weaker evidence (mean evidentiality rating was 62); see Table 3.2 for details.

Table 3.1: Raw counts for each utterance of experiment 1 [2] ordered from most frequent to least frequent.

<table>
<thead>
<tr>
<th>Response</th>
<th>counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>believe</td>
<td>185</td>
</tr>
<tr>
<td>know</td>
<td>171</td>
</tr>
<tr>
<td>think</td>
<td>110</td>
</tr>
<tr>
<td>see</td>
<td>80</td>
</tr>
<tr>
<td>notice</td>
<td>58</td>
</tr>
<tr>
<td>am sure</td>
<td>54</td>
</tr>
<tr>
<td>other</td>
<td>28</td>
</tr>
<tr>
<td>guess</td>
<td>14</td>
</tr>
</tbody>
</table>

5Please be aware, that we excluded guess, and the option other later on e.g. in Figure 3.1 and the analysis. I included the analysis output for a model including guess and other in Appendix 7.1. The results of both analyses are almost identical.
Table 3.2: Mean evidentiality ratings by utterance for experiment 1 [2].

<table>
<thead>
<tr>
<th>Response</th>
<th>mean</th>
<th>sd</th>
<th>median</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>know</td>
<td>81</td>
<td>19</td>
<td>83.0</td>
<td>16</td>
<td>100</td>
</tr>
<tr>
<td>notice</td>
<td>72.3</td>
<td>23.3</td>
<td>71.5</td>
<td>15</td>
<td>100</td>
</tr>
<tr>
<td>am sure</td>
<td>73</td>
<td>22.3</td>
<td>75.5</td>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td>see</td>
<td>79.3</td>
<td>22.2</td>
<td>81.0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>think</td>
<td>62.2</td>
<td>28.2</td>
<td>63.5</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>believe</td>
<td>67.4</td>
<td>22.5</td>
<td>70</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>guess</td>
<td>54.2</td>
<td>18.7</td>
<td>60</td>
<td>15</td>
<td>74</td>
</tr>
<tr>
<td>other</td>
<td>58.5</td>
<td>28.7</td>
<td>69.5</td>
<td>3</td>
<td>100</td>
</tr>
</tbody>
</table>

Contrasting the by-expression evidentiality ratings for both scenarios reveals that, overall, the median evidentiality ratings across expressions were lower for events mentioned in the interrogation than in the briefing; see Figure 3.1. The only exception is *am sure*. If we were to assume that the median evidentiality ratings give insight to the thresholds of the individual expressions, the following order of expression threshold for each scenario emerges:

Briefing:

know > notice > see > believe > am sure > think > guess

Interrogation:

know > see > am sure > notice > believe > think > guess

Compared to the briefing, *see* and *am sure* are much higher ranked in the interrogation, whereas *notice* and *believe* move down in the ranking. The differences in order between the two scenarios would be surprising if we were to strictly assume that speakers choose the expression with the highest possible threshold compatible with their degrees of belief.

For the expressions *notice, am sure, think* and *believe*, an increase in evidentiality meant a decrease in log-odds. This effect was most pronounced for *think* ($\hat{\beta} = -1.13$, CrI: [-1.55, -0.73]), and weakest for *notice* ($\hat{\beta} = -0.38$, CrI: [-0.88, 0.18]), see Table 3.3 for all expressions. This suggests that averaging over scenarios, speakers are less likely to choose *notice, am sure, think* or *believe* than *know* when they are highly certain about the content they want to communicate. The effect was absent for the expression *see* ($\hat{\beta} = 0$, CrI: [-0.41, 0.44]) which suggests that, averaging over scenarios, speakers are free to choose *know* or *see* when they are highly certain.
Figure 3.1: By-expression evidentiality ratings of experiment 1 [2] for each scenario (briefing in blue, interrogation in red). The plot shows the median of the evidentiality ratings (line) and the upper quartile and lower quartile (box). Whiskers extend to the scores outside the the quartiles. Dots represent outliers.

Turning to the effect of scenario, for the expressions *am sure*, *think* and *believe*, the probability to be chosen increased in the briefing and decreased in the interrogation. This effect was most pronounced for *am sure* ($\hat{\beta} =0.84$, CrI:[0.35, 1.39]), followed by *think* ($\hat{\beta} =0.69$, CrI:[0.36, 1.02]) and *believe* ($\hat{\beta} =0.47$, CrI:[0.18, 0.77]). This suggests that, given average evidentiality, speakers choose *am sure*, *think* or *believe* in preference to *know* in the briefing, but less so in the interrogation. The effect was absent for *notice* ($\hat{\beta} =0.18$, CrI:[-0.19, 0.56]) and *see* ($\hat{\beta} =0.12$, CrI:[-0.21, 0.44]). This suggests that given average evidentiality, the extent of speakers’ preference for *know* over *see* and *notice* is independent of scenario.

Overall, the findings suggests that in the briefing, stronger evidence was needed in order for participants to choose expressions associated with a high degree of certainty such as *know*. In the interrogation, however, less convincing evidence might have sufficed. Furthermore, evidentiality was found to play a bigger role than scenario in expression choice.
### Table 3.3: Experiment 1 [2]: 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.

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Coefficient</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>$\hat{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>notice</td>
<td>Intercept</td>
<td>-1.39</td>
<td>0.31</td>
<td>-2.06</td>
<td>-0.84</td>
<td>1.00</td>
</tr>
<tr>
<td>am sure</td>
<td>Intercept</td>
<td>-2.03</td>
<td>0.46</td>
<td>-3.05</td>
<td>-1.26</td>
<td>1.00</td>
</tr>
<tr>
<td>see</td>
<td>Intercept</td>
<td>-0.95</td>
<td>0.24</td>
<td>-1.47</td>
<td>-0.52</td>
<td>1.00</td>
</tr>
<tr>
<td>think</td>
<td>Intercept</td>
<td>-0.70</td>
<td>0.26</td>
<td>-1.25</td>
<td>-0.21</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Intercept</td>
<td>0.02</td>
<td>0.23</td>
<td>-0.43</td>
<td>0.47</td>
<td>1.00</td>
</tr>
<tr>
<td>notice</td>
<td>Evidentiality</td>
<td>-0.38</td>
<td>0.27</td>
<td>-0.88</td>
<td>0.18</td>
<td>1.00</td>
</tr>
<tr>
<td>am sure</td>
<td>Evidentiality</td>
<td>-0.62</td>
<td>0.26</td>
<td>-1.12</td>
<td>-0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>see</td>
<td>Evidentiality</td>
<td>0.00</td>
<td>0.22</td>
<td>-0.41</td>
<td>0.44</td>
<td>1.00</td>
</tr>
<tr>
<td>think</td>
<td>Evidentiality</td>
<td>-1.13</td>
<td>0.20</td>
<td>-1.55</td>
<td>-0.73</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Evidentiality</td>
<td>-0.77</td>
<td>0.21</td>
<td>-1.19</td>
<td>-0.36</td>
<td>1.00</td>
</tr>
<tr>
<td>notice</td>
<td>Scenario1</td>
<td>0.18</td>
<td>0.19</td>
<td>-0.19</td>
<td>0.56</td>
<td>1.00</td>
</tr>
<tr>
<td>am sure</td>
<td>Scenario1</td>
<td>0.84</td>
<td>0.26</td>
<td>0.35</td>
<td>1.39</td>
<td>1.00</td>
</tr>
<tr>
<td>see</td>
<td>Scenario1</td>
<td>0.12</td>
<td>0.17</td>
<td>-0.21</td>
<td>0.44</td>
<td>1.00</td>
</tr>
<tr>
<td>think</td>
<td>Scenario1</td>
<td>0.69</td>
<td>0.17</td>
<td>0.36</td>
<td>1.02</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Scenario1</td>
<td>0.47</td>
<td>0.15</td>
<td>0.18</td>
<td>0.77</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Interim Summary

The results of the first experiment suggest that speakers indeed base their utterance choices on their degrees of belief and moreover adjust their choices depending on the scenario they are in. Overall, participants chose (un)certainty expressions associated with lower degrees of belief more often in the briefing than in the interrogation. For example, think, and believe received lower mean evidentiality ratings than know (62 and 67 respectively, see Table 3.2), and were more likely to be used in the briefing than in the interrogation. There were no differences between scenarios for know and see and the mean evidentiality ratings for see (79.3) were almost as high as the ones for know (81). The results of am sure and notice are less straightforward to interpret: While am sure was more often used in the briefing than in the interrogation, no effect of scenario for notice was found, despite both expressions having received almost the same evidentiality ratings (73 and 72 respectively).

On the face of it, the results of the first experiment suggest a complex pattern of usage preferences without clear indications of the kind of stratification of choices that would be predicted by a threshold semantics account. However, it is possible that this reflects the crowded space of possible
options, and differences of opinion between participants as to the relative strength of, for instance, 

\emph{believe} and \emph{am sure}. Hence, in the following experiment, we constrain the space of choices further in a bid to obtain clearer results. Specifically, we consider \emph{know} in comparison with \emph{believe} across a wide range of confidence levels (\emph{know} expecting to elicit higher, and \emph{believe} lower confidence levels). We use \emph{believe} rather than \emph{think} because \emph{believe} was chosen by participants more frequently.

3.4.2 Experiment 2 [3]: \emph{know} versus \emph{believe}

The second experiment, aimed to investigate the usage of (un)certainty expressions further by focusing on \emph{know} and \emph{believe} based on the results of the first experiment. The experimental design, materials and procedure are the same as for the first experiment with the difference that participants could only choose between \emph{know}, \emph{believe} and ‘other’. Similar to the first experiment the hypothesis is that speakers decide between \emph{know} and the epistemically weaker option \emph{believe}, depending on (i) their degree of confidence, (ii) and the communicative setting they are in.

Participants

85 participants were tested and recruited over the crowd-sourcing platform Prolific, specifying participants with an approval rate above 90 and restricted to people that had not previously participated in experiment 1 [2]. Participants were paid with an average of £7.53/hour (the average duration of the experiment was 30 minutes). After data collection two participants were excluded because their accuracy for the control items was below chance level. The age of the remaining 83 participants ranged from 18 to 66 years, with a mean of 32 years. 50 participants stated their preferred pronoun as she/her, 32 chose he/him and 1 chose they/their.

Materials

The materials were the same as for the first experiment, see section 4.5.1. The only difference was that the utterance choice was limited to \emph{believe} and \emph{know}, see (3.4.2) as an example for the Briefing scenario.

\begin{enumerate}
\item Briefing item:
\begin{enumerate}
\item Participant: Did Emily Brown have any financial problems?
\item Colleague: Financially the suspect was doing alright.
\item Participant: I _______ that the suspect was in need of money.
\end{enumerate}
\end{enumerate}

[\emph{know}|\emph{believe}]
Procedure

The procedure matched that from the first experiment, see section 4.5.2.

Analysis

The data was analysed fitting a Bayesian binary logistic regression model with maximal random effects structure using the R (R Core Team, 2020) package brms (Bürkner, 2018). The experimental factor scenario (briefing/interrogation) and the continuous variable evidentiality ([0, 100]), were included to predict the probability to choose know. The factor scenario was again sum-coded (-1 as interrogation and 1 as briefing) and evidentiality was standardised. The model included varying intercepts and slopes for participants and items.

The same priors as for the data of the first experiment were used: The intercept was normally distributed with mean 0 and standard deviation 10. For the fixed effects, normal priors with a mean of 0 and a standard deviation of 1 were used. The standard deviations were assigned half-normal priors with a mean of 0, and a standard deviation of 1 and for the correlation matrix, a LKJ(2) prior was used.

The sampling process was the same as for the experiment 1 analysis, see section 4.5.2 for a detailed description. The response ‘other’ was again excluded from the analysis. As anticipated, providing participants with a reduced set of utterances, ‘other’ was chosen more frequently than in the first experiment and made up 19% of the data. After examining the data, ‘other’ was often chosen when evidentiality ratings were comparably low. We will leave the analysis of these responses for future research but will not focus on them in this study.

Results

Two participants were excluded for below-chance performance on the control items. The remaining participants responded to the control items with an accuracy of 97%. Similar to the first experiment, in the evaluation task, participants assessed the evidence by using the whole range of the slider: the evidentiality measure ranged from 0 to 100, with a mean of 72. Furthermore, similar to the first experiment, the means, standard deviations and medians of the evidentiality measures for the scenarios briefing and interrogation were similar (mean=72.7/71.7, sd=26.7/27.9, median=80/79 respectively). This suggests that evidentiality did not vary meaningful between scenarios.

Note that here the option other was included which was later excluded for the more detailed analysis. We carried out a further analysis which included other in Appendix 7.1. The results of the main model excluding other and the model including other were fairly similar considering that we are comparing a binary logistic regression model with a categorical logistic regression model.
Figure 3.2: By-expression evidentiality ratings of experiment 2 [3] for each scenario (briefing in blue, interrogation in red). The plot shows the median of the evidentiality ratings (line) and the upper quartile and lower quartile (box). Whiskers extend to the scores outside the the quartiles. Dots represent outliers.

<table>
<thead>
<tr>
<th>Response</th>
<th>mean</th>
<th>sd</th>
<th>median</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>know</td>
<td>83.45</td>
<td>27.12</td>
<td>91</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>believe</td>
<td>64.67</td>
<td>27.12</td>
<td>70</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 3.4: Mean evidentiality ratings by utterance for experiment 2 [3].

With regards to differences between formulations, on average know (mean 83) was chosen for higher evidentiality measures than believe (mean 65), see Table 3.4 for more information. Looking at the by-expression evidentiality ratings contrasting both scenarios, it seems that the evidentiality ratings across expressions were lower for events mentioned in the interrogation than in the briefing; see Figure 3.2. The difference seems to be more pronounced for believe. This suggests that in the briefing, stronger evidence was needed in order for participants to choose know. In the interrogation, however, less convincing evidence might have sufficed.

These observations are supported by the outcome of the Bayesian analysis. The estimate of the main effect of evidentiality was ($\hat{\beta} =1.34$, CrI:[1.02, 1.70]), suggesting that the probability of saying know increases when the speakers’ degrees of belief increase, see Table 3.5 for more details. The
estimate of the scenario effect was ($\hat{\beta} = -0.29$, CrI: [-0.50, -0.10]), suggesting that speakers were more likely to say *know* in the Interrogation scenario than in the Briefing scenario. Thus, the effects found in the first experiment could be replicated.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>$\hat{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.48</td>
<td>0.19</td>
<td>-0.87</td>
<td>-0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>Evidentiality</td>
<td>1.34</td>
<td>0.17</td>
<td>1.02</td>
<td>1.70</td>
<td>1.00</td>
</tr>
<tr>
<td>Scenario</td>
<td>-0.29</td>
<td>0.10</td>
<td>-0.50</td>
<td>-0.10</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 3.5: Experiment 2 [3]: Population-level estimates of the categorical regression model in log-odds with the standard errors and 95% credible intervals. 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.

We visualised the model’s predictions given the data in Figure 3.3. In order to facilitate understanding we back-transformed the data from log-odds to probabilities. The plot shows the predicted probabilities of using *know* for each scenario. The x-axis represents the standardised evidentiality ratings, whereby 0 means average evidentiality (corresponding to 73 on the original scale). The lines represent the means of the fixed effects and the faded area the 95% credible intervals for the effects. The plot illustrates that with increasing evidentiality, the probability to use *know* is predicted to increase as well. Furthermore, *know* is predicted to be more likely to be used in the Interrogation scenario than in the Briefing scenario. For example, with an average evidentiality rating (0 in the plot), the probability of using *know* is predicted to be approximately 0.31 in the briefing and approximately 0.45 in the interrogation.

However, inspecting credible intervals of the posterior means, we can see that the estimated distribution for evidentiality is further away from 0 than the estimated distribution of scenario. This suggests that evidentiality plays a bigger role for speakers deciding between *know* and *believe* than the scenario they are in.

**Discussion**

The findings of the second experiment replicated the findings from the first experiment: speakers seem to base their utterance choices on their degrees of believe and moreover to adjust their choices depending on the scenario they are in. For experiment 2 [3] this means concretely that participants chose *believe* for lower evidentiality ratings than *know* and that they were more likely to choose *believe* in the briefing than in the interrogation. One difference between the two experiments lies in the evidentiality ratings for *believe* which were lower for the second experiment than for the first. It
Figure 3.3: Experiment 2 [3]: Predictions of the model given the data. Log-odds were back-transformed to probabilities (y-axis). The x-axis is the standardised evidentiality measure: 0 stands for an evidentiality of 73. An increase of one standard deviation on the standardised scale means an increase of 26 on the original scale. The lines represent the means of the fixed effects and the faded area depicts the 95% credible interval of the fixed effects.

is reasonable to assume that participants were willing to use believe for lower degrees of confidence in the second experiment due to the lack of explicitly suggested alternatives.

3.5 General Discussion

There are two different ways to interpret these findings. First, they could suggest that in a cooperative scenario speakers want to truthfully communicate their certainty in order to be as informative – select the verb with the highest threshold that does not exceed their degree of belief in the complement proposition – and cooperative as possible. In such a scenario a speaker chooses a particular expression based on their degrees of belief such that, if they deem the event probability to exceed the threshold of an expression, the speaker chooses that expression as long as there is no other,
more informative expression. In contrast, in a scenario where speakers are faced with an uncooperative interlocutor who might not tell the truth, speakers might instead act strategically, and use expressions that are associated with higher certainties such as know or see in order to come across as authoritative and confident (Hosman, 1989; Crismore & Kopple, 1997).

Alternatively, the findings could suggest that in a cooperative scenario speakers are obliged to engage to some extent in polite conversation. This would mean that, besides pursuing the aim to be cooperative and most informative, speakers also engage in facework. Hence, a briefing with a colleague would not only entail acting cooperatively and conveying information to each other but also trying not to step on each other’s toes. Recall that, in hedging accounts and politeness theory (e.g., Fraser, 1975; Brown & Levinson, 1987), this means that cooperative speakers tend to down-tone their utterances when communicating their certainty in order to be polite. By contrast, in an uncooperative scenario speakers might feel less obliged to engage in facework especially when they are in a high-power position. Thus, instead of hedging their statements, speakers communicate their degrees of belief directly.

While we incline towards the first explanation, we acknowledge that the second explanation is hard to rule out experimentally, given the putatively ubiquitous nature of facework. Although the Briefing scenario is designed to be cooperative, in the sense of prioritising the exchange of accurate information, politeness considerations are doubtless still at play to some extent. These considerations may mediate between the speaker’s degree of belief and choice of (un)certainty expression in a way that has yet to be fully theorised. This is due to the fact that there is no traditional baseline condition. We purposely chose a detective story as a cover story because it enabled us to create a coherent experiment where being part of a cooperative and an uncooperative scenario is plausible for the participant. Even in a more neutral context, such as talking to a friend, conventions such as being polite will be involved to some extent so it is unclear whether this would constitute a suitable baseline condition.
Figure 3.4: Experiment 2 [3]: The plots show the by-participant usage of *know* (purple dots) and *believe* (green dots) in the two different scenarios. The y-axis represents the individual participants, whereby each grey line belongs to one participant. Participants are ordered from top to bottom according to the lowest degree of belief for which they used *know*. The x-axis represent the evidentiality measure. This way we can see the degrees of belief of the participants when choosing a particular expression. A green dot that appears immediately to the right of a purple dot indicates a case where a speaker used *believe* for stronger evidence than for which they elsewhere used *know*.
In practice, the results suggest that speakers use (un)certainty expressions much more dynamically than expected by a strict threshold semantics account, if that account is coupled with standard pragmatic assumptions about cooperativity. For example, we did not find a fixed ordering of (un)certainty expressions. Recall that, under these assumptions, we would expect each expression to be restricted to evidentiality levels between its threshold and the threshold for the next stronger expression. For instance, given the choices of *think* and *know*, we would expect to see *think* attested for evidentiality above the threshold for *think* and below the threshold for *know*. However, we did not find a fixed ordering of the kind this model would predict. Considering the median evidentiality rankings for each (un)certainty expression from experiment 1 [2], repeated below, we found differences in their ordering between scenarios.

Briefing:

know > notice > see > believe > am sure > think

Interrogation:

know > see > am sure > notice > believe > think

As the statistical analysis above suggests, many participants appeared to be willing to use *know* for lower evidentiality ratings in the Interrogation scenario than in the Briefing scenario. This could be accommodated within a threshold-based account by assuming that an individual’s threshold for the use of *know* can vary between scenarios, just as thresholds are argued to vary between speakers in general. However, a more challenging result for the threshold-based account is that participants frequently used *believe* for levels of evidentiality which exceeded those for which they elsewhere used *know*. This is shown in Figure 3.4 where we plotted the by-subject utterance choices across evidentiality ratings.

These findings could be reconciled with the threshold-based account in a couple of ways. One possibility is to assume that an individual’s threshold for using *know* varies throughout the experiment (or that the participants’ reported evidentiality ratings do not correspond with the ratings on which they based their productions at the time of utterance). Another possibility is that speakers are simply not pragmatic in the way we have assumed, and freely produce less informative utterances than they are entitled to (i.e. asserting *believe* in just some specific circumstances when they could assert *know*), or produce more informative utterances than are warranted (i.e. asserting ‘know’ when they do not have sufficient confidence, by their own criteria, to license this), which is likely given that speakers do not behave perfectly optimal or informative. However, within this variability we have to assume some sort of pragmatic behaviour within a given theory, otherwise the claims of
that theory are difficult to falsify.\footnote{In Chapter \textbf{4} Section \textbf{4.6.3}, we go more into detail, how threshold semantics could capture our data.}

In terms of the broader implications for theories of communication, the results suggest that speakers choose an utterance not only based on their perception of the world (here their degrees of belief) but also on the effect that their communicative action may have on the hearer. Sperber and Wilson (1986/1995) claim that speakers’ intention is twofold: speakers (i) want to be understood (informative intention), and (ii) aim to convince their hearers to think or act in accordance with the speakers’ beliefs (communicative intention). In particular, the results suggest that speakers are willing to make stronger claims than appear to be semantically warranted, in a context in which doing so could be communicatively effective in the latter sense.

This in turn raises the question of how a hearer should act in such a case – that is, where they suspect that a speaker is overstating the probability that a proposition is true, according to the usual assumptions about the meanings of (un)certainty expressions. Sperber et al. (2010) suggested that interlocutors automatically engage an ‘epistemic vigilance’ mechanism whose purpose is to detect misinformation, by assessing the quality and plausibility of the communicated content as well as the reliability of the speaker. In the Interrogation scenario, it would be rational for a hearer to engage in an even more conscious evaluation of credibility – that might go beyond the automatic application of the epistemic vigilance mechanism – and opt out of the usual cooperative assumption that the speaker’s claims are to be accepted at face value.

Research on epistemic vigilance (e.g., Mazzarella, Trouche, Mercier & Noveck, 2018) has typically focused on cases where a proposition is categorically accepted or rejected, so it is potentially interesting to consider what epistemic vigilance means in the context of gradable degrees of belief. Confronted with a claim that the speaker ‘knows’ p, should the vigilant hearer conclude that the speaker in fact merely ‘believes’ p, or should they reject it outright?

Moreover, keeping in mind that hearers are vigilant, speakers have to be cautious and check to what extent their utterance coheres with the beliefs of the hearer in order to make it more probable for their communicative intention to succeed (Sperber et al., 2010). In this experiment, participants might have reasoned similarly following an overall goal of gaining compliance (e.g Dillard, 1990). In the Briefing scenario, for example, they might have deemed it strategically advantageous to occasionally communicate their degrees of belief truthfully without under- or overstating their confidence. Here, the aim of truthfully communicating their confidence would be to prove their reliability which could facilitate their overall goal: briefing their colleague successfully. Similarly,
in the Interrogation scenario, the participant might wish to be perceived as reliable, in order to avoid the suspect engaging in epistemic vigilance to an enhanced degree and thus discounting subsequent overstated claims. Thus, throughout the experiment, there could be higher-order strategies in play, which might explain participants’ varied use of (un)certainty expressions within one scenario – the speaker may wish to give the hearer a particular impression of how they use (un)certainty expressions, much like how a good poker player will fold on some bad hands in order to make a subsequent bluff more effective. This could explain participants’ varied usage of (un)certainty expressions within one scenario. However, these claims have to be tested further.

Moreover, it remains to be investigated whether speakers can employ these strategies successfully, especially in the interrogation setting where the speakers’ goal is to convince the hearer and the hearer has a great incentive to hide the truth. Future research may also shed light on what parts of the communicated content hearers might refuse. Overall, the study illustrates participants’ expectations of what someone else would say in a given situation, and as such, indirectly gives insights into the comprehension of (un)certainty expressions. In the end, speakers and hearers are both interlocutors who inevitably reason about each other and constantly exchange their roles within their dialogue.

In the broader context of uncertainty, we could see these interactions as attempts to reduce the interlocutors’ uncertainty about the truth-values of various propositions under discussion. From this point of view, a police officer’s interrogation of a suspect could be seen as an attempt to eliminate their subjective cognitive uncertainty as to whether the suspect is guilty. The use of an expression that conveys a false degree of confidence in the factuality of a proposition – e.g. ‘I know you were there’ – could be one way of seeking confirmation of a hitherto uncertain claim, and thus building a common ground as to what is true. The briefing of a colleague, by contrast, does not aim at the complete elimination of uncertainty, but towards the building of a consensus between the interlocutors as to how (un)certain various propositions subjectively should be, given the available evidence. This difference potentially underlies the difference in communicative strategies evident in these experiments.

3.6 Conclusion

In this experiment, we tested the production of (un)certainty expressions in two contrasting scenarios. The first experiment contrasted a wide range of (un)certainty expressions, and the second experiment focused on the production of know versus believe. We found that speakers choose between (un)certainty expressions based on their degrees of belief, and furthermore, adjust their
choices depending on the scenario they are in. These findings support hedging and politeness accounts which assume that speakers may use (un)certainty expressions strategically. By contrast, our findings are surprising under a strict threshold semantics account paired with pragmatic assumptions about cooperativity, since we found that speakers use (un)certainty expressions much more freely. Besides using *know* for lower evidentiality ratings in the Interrogation scenario than in the Briefing scenario, participants frequently used *believe* for levels of evidentiality which exceeded those for *know*.

### 3.7 What’s next?

Experiments [2–3] shed light on the production of a variety of (un)certainty expressions, the factive presupposition trigger *know* and the degrees of belief they convey. The results suggest that speakers indeed base their utterance choices on their degrees of belief, and furthermore, adjust these choices depending on the communicative setting they are in. However, the results of experiments [2–3] are compatible with two different interpretations: (i) speakers utter *know* more readily in uncooperative settings and thereby overstate their knowledge, or (ii) speakers utter *believe* more readily in cooperative settings and thereby understate their knowledge in order to come across as polite. Thus, both expressions may be used strategically with opposite effects.

So far we have investigated the production of *know* and its potential strategic advantage over (un)certainty expressions on the level of conveyed degrees of belief: because *know* conveys the highest degrees of belief its usage may be strategically advantageous in uncooperative settings. For example, speakers who are faced with a potentially lying interlocutor might want to overstate their knowledge and thereby provoke the interlocutor to tell the truth. However, this might oversimplify the production of *know*. Being a presupposition trigger, *know* might be a useful strategic tool for speakers to influence what content gets picked up in the subsequent discussion.

In order to shed more light on the discourse effects of *believe* and *know*, Part II proceeds by focusing on the distinction between uttering *believe that p*, *know that p* and simply saying *p* (i.e. the bare assertion). For the distinction between *believe* and the bare assertion, we would like to see whether speakers favour uttering *believe that p* over *p* in cooperative settings and thereby hedge the assertive strength of *p* to come across as polite. For the distinction between *know* and the bare assertion we aim to investigate whether speakers would favour presupposing *p* (by uttering *know that p*) over asserting *p* in uncooperative settings, with the intention of influencing what content is being picked up in the subsequent discourse.
Chapter 4

Strategic reasoning when formulating and comprehending knowledge ascriptions

4.1 Abstract

In communicating about certainty, speakers make choices among available formulations and hearers will aim to recover speaker intentions. In two studies [experiments 4 and 5], we assess speakers’ production choices and hearers’ interpretations to test (a) how maximal certainty is formulated, (b) whether those formulations adjust depending on context, and (c) whether speakers’ context-driven adjustments are apparent to hearers. We compare the lower-certainty formulation I believe that the deadline is tomorrow [believe] with two high-certainty formulations, I know that the deadline is tomorrow [know] and The deadline is tomorrow [bare assertion]. Given debates about which of the high-certainty formulations should convey higher epistemic standards, we investigate when (if ever) know is favoured, perhaps for its felicitous use in a wider range of contexts or perhaps for its strategic use in structuring subsequent dialogue via presupposition and accommodation. We investigated whether interlocutors align in the way they convey and recover meaning from statements about degrees of belief, comparing their behaviour across cooperative and uncooperative scenarios. Our results suggest: (a) know is favoured for maximal certainty, followed by the bare assertion and believe for content with successively lower degrees of belief; (b) speakers make context-dependent adjustments whereby they use know strategically in uncooperative scenarios to overstate their knowledge. However, with regards to (c) we did not find that hearers recover these production strategies. The misalignment between speakers and hearers may reflect our measure of comprehension from a bystander point of view, but it might also indicate that speakers succeed with their strategic
4.2 Introduction

Besides introducing propositional content, speakers are often interested in conveying their sense of how likely the proposition is to be true. This gives the hearer potentially valuable information both about the state of the world and the knowledge state of the speaker. In communicating their level of certainty regarding propositional content, speakers make choices between available formulations, which include directly asserting content, or using a sentential verb such as believe or know. These competing formulations convey different degrees of belief: believe is associated with relatively low degrees of belief and in that way differs from know and the bare assertion. However, it is much less clear in which way the two high-certainty formulations compete with each other. There have been a variety of epistemic approaches but most of the arguments that have been put forth have focused on the felicitous usage of know or on the impact of assertion, so that the question about how to systematically compare them remains open. Besides potentially conveying different degrees of belief, know and the bare assertion may differ in the extent to which they put forward propositional content and thereby affect the proceeding discourse structure: Speakers uttering know presuppose content and treat it as already shared information that does not require further discussion; in contrast, a speaker uttering the bare assertion asserts content which is generally up for discussion.

In this study we would like to know under which circumstances (if any) speakers prefer to utter know over the bare assertion. For this purpose, we will examine the relationship between sentences with know and the corresponding bare assertions, taking into account both epistemic considerations and the implications for dialogue structure. More specifically, this paper will examine speakers’ production choices and hearers’ interpretation of the formulations believe/know/bare assertion to test (i) how maximal certainty is expressed and inferred; and (ii) whether speakers adjust their production choices/the way they introduce content depending on their communicative goals and 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 [experiment 4] and comprehension [experiment 5] respectively. We will first introduce theoretical accounts and their predictions which we will then discuss in the light of our production and comprehension results.
4.3 Epistemic Strength

4.3.1 The Knowledge Norm of Assertion and Contextualism

The epistemic accounts that we will introduce discuss the epistemic prerequisites for the bare assertion and \textit{know}. These accounts are conflicting and often only pertain to \textit{know} or the bare assertion, whereas the goal of this paper is to compare the two. We discuss and compare these accounts to highlight the reasons why either form can be seen as epistemically strong. The upshot, we argue, is that additional empirical evidence is needed to inform these theories and to understand to what extent speakers meet the claimed epistemic prerequisites when uttering these two forms.

Although both \textit{know} and the bare assertion are typically uttered to convey high degrees of belief, competing accounts posit different epistemic prerequisites for their use. As an example, we will consider the proposition \( p = \text{The major is corrupt} \) where (39) asserts \( p \) and (40) places \( p \) in the scope of \textit{know}.

\begin{align*}
(39) & \quad \text{The major is corrupt.} \\
(40) & \quad \text{I know that the major is corrupt.}
\end{align*}

For Williamson (2000), only knowledge warrants assertion, and hence, one can truthfully assert \( p \) only if one knows \( p \). This claim is based on Moore’s (1942) paradox and the observation that ‘\( p \) but I don’t know that \( p \)’ is incoherent, as well as Unger (1975)’s argument that “How do you know?”, which presupposes that the speaker knows \( p \), is a felicitous response to an assertion of \( p \). It has to be noted that speakers in casual conversation do not have to ‘know’ to a very high evidence standard in order to assert information. However, since the nature of the scenarios (briefing/interrogation) used in the experiments require a high standard of evidence we will be working with predictions based on what we have introduced so far.

The Knowledge Norm of Assertion was targeted at the prerequisites for assertions (\( p \)) and does not discuss differences between the prerequisites of \( p \) and \textit{know that} \( p \). One could conclude that this account extends to \textit{know}, since the bare assertion and \textit{know} are logically equivalent: if one is able to assert \( p \), since one knows \( p \) one should similarly be able to utter \textit{know that} \( p \). We will focus only on the prerequisites for assertion for now. Furthermore, it has to be emphasised that we take the Knowledge Norm of Assertion to be solely about the prerequisites of assertion not about speakers’ actual production preferences between uttering \( p \) or making their knowledge claim explicit (\textit{know that} \( p \)).
One objection to this form of 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 Grice (1975) intended 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 (according to what we have introduced here as the Knowledge Norm of Assertion) being in the position to know that p, utterances about the future, e.g., ‘the sun will come up tomorrow’ would be rendered infelicitous.

So far, knowledge has been discussed in absolute terms. This contrasts with Epistemic Contextualism which has focused principally on knowledge claims such as I know that p whereby knowledge is seen as gradable. Here, the epistemic prerequisites of uttering I know that p depend on the context in which they are uttered (e.g. DeRose, 1992; Cohen, 1999). More specifically, the epistemic standards of a speaker for uttering I know that p may vary. For example, a courtroom context may impose different epistemic standards on a speaker than a classroom setting. DeRose (1992)’s Bank Case, here repeated from Stanley (2004) as (41), illustrates this point:

(41) Hannah and her husband are driving home on a Friday afternoon. They plan to stop at the bank on the way home to deposit their paychecks. But as they drive past the bank, they notice that the lines inside are very long, as they often are on Friday afternoons. Thinking that it isn’t very important that their paychecks are deposited right away, Hannah says “I know the bank will be open tomorrow, since I was there just two weeks ago on Saturday morning. So we can deposit them tomorrow morning.” But then Hannah’s husband reminds her that a very important bill comes due on Monday, and that they have to have enough money in their account to cover it. He says, “Banks do change their hours. Are you certain that’s not what is going to happen tomorrow?” Hannah concedes, uttering “I guess I don’t really know that the bank will be open tomorrow.”

This example suggests that the propositions expressed by the same speaker differ in respect to these two different contexts. In the context of Hannah not being aware of the importance of depositing the money tomorrow, utterance (i) I know that p seems to convey that p is true. Then when Hannah’s husband reminds her of the high stakes involved, the proposition expressed by the utterance (ii) I guess I don’t really know that the bank is open tomorrow morning equally seems to be true. DeRose (1992) argues that (ii) is not the denial of (i) because know functions as a context-sensitive term analogous to indexical expressions. Relative to different contexts of use, it expresses different relations between persons and propositions. So know has a different content in utterance (i) than in (ii). However, (41) remains felicitous if Hannah had uttered the bare assertion The bank will be open tomorrow, since I was there just two weeks ago on Saturday morning. for (i). This would suggest that it is not solely about know having different contents in utterances (i) and (ii) and that
(41) is merely an example of how speakers can redeem themselves when being called out on having expressed too much confidence, namely by uttering hedging device *I guess* and to deny that they know p in conjunction with *really*.

By understanding knowledge as gradable, contextualists similarly understand *know* to be a gradable term. However, this claim has been challenged. *Know* does not allow for degree modifiers (#I very much know that p / #I strongly know that p) whereas *believe* can be combined with (I very much believe that p/ I strongly believe that p; Stanley (2004)). Yet if one were to understand knowledge as justified true belief, the usage of a small set of degree modifiers is possible (Partee, 2004). For example (taken from Partee (2004)), (42) relates to the degree of certainty/belief and in (43) relates to the degree of justification.¹

(42) Hannah knows without a doubt that p
(43) Hannah knows conclusively that p, knows reliably that p

According to the Knowledge Norm of Assertion, speakers who want to assert p have to know that p is true. Turning away from the bare assertion, in Epistemic Contextualism, *know* is seen as context-dependent: speakers can felicitously utter *know that p* as long as the epistemic standards are met which are dictated by the context and may be sometimes lower or higher. Taking both accounts together suggests that *know* is epistemically weaker than the bare assertion, since *know* may be uttered in contexts with a more diverse range of epistemic standards than the bare assertion.

That it is instead *know* that conveys higher degrees of belief is assumed by a later proposal of contextualism which we will call the Knowledge in Context account. DeRose (2002) claims that given the Knowledge Norm of Assertion – one is positioned to assert p only if one knows p – one can assume that a speaker who is in the epistemic position to assert p should be similarly in the position to utter *I know that p*. Both formulations are understood to be context-dependent, i.e. both *know p* and p depend for their felicitous usage on the speaker meeting the epistemic requirements for the given context - that is to say, that the argument from DeRose’s Bank Case (repeated as (41) in the paper). This is based on DeRose (2002)’s claim that since both utterances are subject to

¹Note that, independently of whether *know* is gradable or not, the take-away hypothesis for contextualism is that the epistemic requirements to utter *know* are understood to vary by context. If one were to deny that *know* is gradable, the variation between contexts could be understood as a variation in precision, following Lewis (1979). For example, in (1) the non-gradable formulations *hexagonal* and *boot-shaped* are uttered with a certain amount of slack and seem felicitous given a lax context.

(1) a. Italy is boot-shaped.
   b. France is hexagonal.
epistemic standards per se – the standards of knowledge – if the standards for know vary so should the standards for the bare assertion. Crucially, DeRose (2002) argues that I know that p has higher epistemic standards than the bare assertion and that know is therefore epistemically stronger. This contrasts with what was suggested by the Knowledge Norm of Assertion about the bare assertion and Epistemic Contextualism about know. This argument reflects a distinction between what is being asserted in these two cases (p versus knows that p) and what the two assertions represent respectively (the speaker signals that they know that p versus the speaker signals that they know that they know that p). Thus, according to the Knowledge in Context Account know may be preferred over the bare assertion to express maximal certainty. In the next section, we will focus on empirical evidence on the production of the bare assertion and know.

4.3.2 Empirical Findings

Empirical evidence supports the intuition that the bare assertion is only used when the speaker is highly certain about the propositional content. In a study by Schuster and Degen (2020), participants were found to choose the bare assertion when the event that its proposition described had a probability very close or equal to 1. Among the available choices in their study were uncertainty expressions such as probably, might and think. Similarly, findings from Turri (2010) suggest that strong evidence alone is not enough for speakers to choose the bare assertion, which they reserve for cases where the proposition in question is unequivocally true.

However, Degen, Trotzke, Scontras, Wittenberg and Goodman (2019) obtained contrasting findings. In a series of experiments, they first assessed the strength of evidence that was presented in the form of propositions by asking participants to evaluate the proposition by adjusting a slider ranging from impossible to absolutely certain. Then, in a following production experiment, they investigated when speakers use evidential devices such as the bare assertion, must, probably and might for the previously rated propositions. Their findings suggest that the bare assertion can be used more freely and for lower degrees of belief than was suggested by Schuster and Degen (2020). The differences between these experiments is potentially traceable to the different way of presenting event probabilities to the participants (quantitative\(^2\) versus qualitative prompts\(^3\)). However, it is unclear which of these approaches is more ecologically valid.

Turning to know, some empirical evidence suggests that know is to some extent gradable and

\(^2\)Schuster and Degen (2020) showed participants an urn filled with blue and red balls with varying ratios and asked them how likely it is to draw a blue ball. This is what we will refer to as a quantitative prompt, since the participants’ estimation of how likely an event is relies on mathematical/statistical/logical knowledge.

\(^3\)We understand a qualitative prompt to be a prompt that does not involve statistical thinking but subjective assessment, e.g. assessing how likely an event is based on a piece of evidence.
context-sensitive. Lorson, Cummins and Rohde (2021)’s (Chapter 3) and Tonhauser, Beaver and Degen (2018)’s studies showed that speakers uttered *know* for a wide range of degrees of belief which suggests that *know* indeed is gradable, albeit to a lesser extent than *think* or *believe*. Furthermore, Lorson et al. (2021) (Chapter 3) found that the usage of *know* depended on the scenario participants were in. More specifically, participants restricted their use of *know* to higher degrees of belief in fully cooperative settings, but were more flexible in less cooperative settings. However, both studies only compared *know* with other verbs that take sentential complements but not with the bare assertion.

Neither the theoretical accounts nor the empirical findings discussed above seem to paint a clear picture about the epistemic prerequisites for either the bare assertion or *know*. Furthermore, we know of no empirical studies that have focused on both expressions together even though such studies would illuminate the extent to which these two formulations differ epistemically from each other. Thus, it remains unclear whether the speaker’s choice between the two formulations is conditioned by considerations of intrinsic epistemic strength, contextual factors, or some combination of the two. In the following section, we explore a further factor that could influence the choice, namely the consequences of that choice for the ongoing discourse.

4.4 Are factives a communicative strategy?

4.4.1 The Presupposition Strength Account

Apart from communicating potentially different degrees of belief, *know* and the bare assertion may differ from each other with respect to their discourse effects. The discourse effects of bare assertions have been contrasted in the literature with uncertainty expressions such as *think/believe, might* and *probably*. Whereas a speaker who asserts p proposes to add p to the common ground, uncertainty expressions can be used as *downtoners* (Holmes, 1982) or *hedges* (Lakoff, 1973; Fraser, 1975; Brown & Levinson, 1978) to weaken the assertive strength of p, yielding discourse effects such as politeness. For example, instead of straight out asserting p (44a), a speaker might want to hedge the assertive strength (44b) to come across as more polite.⁴

(44) a. Your soup needs more salt.

  b. I think your soup needs more salt.

⁴Using *think* weakens the assertive strength of *Your soup needs more salt* and signals that the speaker is not as committed to this claim as if they would have asserted the same. Using *think* signals that whether adding more salt or not is still up for discussion. In contrast, flat out asserting the same claim may signal that there is not question of whether salt should be added or not (it should).
Here, we will be focusing on the potentially important distinction between asserting and presupposing and the corresponding discourse effects. In work on common ground and discourse updating, *know* is understood to trigger a factive presupposition. In (45–46), the speaker presupposes that the deadline is tomorrow. This content can be identified as a presupposition of (45) and (46) on the basis of its ability to project from under the scope of operators such as negation: despite (46) 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.

(45) She knows that the deadline is tomorrow.
(46) 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 Fintel (2008)](#), hearers who lack the shared knowledge are argued to 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 challenging it. This generally holds for cooperative communication, in which presupposed content is sincerely taken by the speaker to be uncontroversial.

One potential reason why a speaker would choose to presuppose rather than assert content when introducing new information is that presupposing is often more concise than explicitly asserting the corresponding information. For example, in the case of (47a), the expression *my car* carries an existential presupposition about the speaker having a car. Asserting the same content would be much less concise/more wordy, as in (47b).

(47) a. Sorry I’m late: my car broke down.
   b. Sorry I’m late: I have a car and it broke down.

At least in terms of the surface form (referring to sentence length), the use of *know* as a presupposition trigger does not appear to offer the same advantages: (48a) is similar in length to (48b), which asserts the proposition of interest (anaphorically) rather than presupposing it.\(^5\)

(48) a. I know that you went to the party anaphorically rather than presupposing it.

\(^5\)An anonymous reviewer suggested that the notion of conciseness may be extended to the dimension of discourse moves: (47b)/(48b) introduce two assertions which requires the hearer to update their common ground twice. In contrast, (47a)/(48a) only introduce one assertion requiring the update of one proposition. However, this argumentation only holds if accommodating a presupposition does not count as a separate discourse move in its own right.
b. You went to the party without me and I know it.

Another consideration for speakers when choosing between asserted and presupposed content could be the degree to which they want to commit to content. We understand speaker commitment to be different from signalling the degree to which speakers believe a proposition to be true, i.e. communicating the speakers’ epistemic state. When asserting content speakers commit themselves to its truth (e.g. Brandom, 1983). The more committed to content a speaker signals they are, the more they endorse this piece of information. This comes at the cost of being more accountable as a speaker and more likely to be subject to scrutiny. Focusing on the distinction between the production of know and the bare assertion, it is not entirely clear whether presupposing information is more or less committal than asserting that same information.

Based on the characteristics of presuppositions being backgrounded, uncontroversial and already shared information, one may argue that speakers commit less to presupposed content and may be less accountable for what they presupposed because it has supposedly already been accepted by all interlocutors. On the other hand, presupposed content may represent material that is well accepted, whereas asserted content could represent the material that the speaker knows is up for debate since it is being proffered but could be refuted. Indeed, recent experimental work by Mazzarella, Reinecke, Noveck and Mercier (2018) found no evidence of a difference in perceived speaker commitment between assertion and presupposition. Their study contrasted IMPLICATING, PRESUPPOSING and ASSERTING by investigating a wide range of presupposition triggers (not including factive verbs). Participants judged speakers to be less blameworthy when they implicate rather than presuppose or assert false information. Overall, Mazzarella, Reinecke et al. (2018)’s findings suggest that IMPLICATING is less committal than PRESUPPOSING and ASSERTING but found no significant difference between the latter two.

However, in this paper we will focus on still another reason why speakers might choose to utter presupposition triggers – namely that they wish to signal what content should be picked up in the subsequent dialogue. As was mentioned before, presupposed content is typically not expected to be discussed further, since it conveys supposedly already shared knowledge. In keeping with this intuition, presupposed information, unlike asserted information, requires additional effort to address in the subsequent discourse. For example, if a speaker would utter (49b) in response to (49a), they are most often taken to deny the assertion (that the speaker’s car broke down) rather than the presupposition (that someone has a car). This has to do with projection: if we simply negate (49a) the presupposition (that someone has a car) still projects.

(49) a. Speaker1: “Did your car break down?”
b. Speaker2: “That’s not true!”/“No!”

The ‘Hey, wait a minute test’, proposed by von Fintel (2008) after Shannon (1976), is a diagnostic for presuppositions that exploits this observation. If speakers aim to deny a presupposition they have to use a circumlocution such as ‘Hey wait a minute...’, as in (50), which is a dispreferred means to deny asserted content, as shown in (51). This corroborates the claim that addressing presupposed content is more effortful than addressing asserted content.

(50) Hey wait a minute, I don’t have a car!
(51) ?Hey wait a minute, my car didn’t break down!

Assuming a relative lack of addressability of presupposed content, one may hypothesise that cooperative speakers should only presuppose information that is uncontroversial and avoid sneaking in controversial information in a way that makes it difficult for hearers to argue with (Schwarz, 2019). Similarly, on the hearers’ side, new information that has been presupposed can be accommodated by default based on the assumption that it is most likely uncontroversial material. Thus, hearers might be more inclined to accommodate and less inclined to challenge presupposed than asserted content.

This tendency may be exploited by speakers who aim to maximise the chances of the hearer accepting the proposition without further discussion. More specifically, by presupposing rather than asserting information, speakers would act as if (potentially controversial) material was already accepted as part of the common ground and signal that they do not consider it to need further discussion. Note that speakers do not necessarily have to possess maximal certainty or commitment in order to pursue this communicative strategy. Nonetheless, this strategic advantage of know over the bare assertion in putting forward propositions may be the reason as to why speakers would choose know over the bare assertion in specific contexts.67

---

6 An anonymous reviewer pointed out that speakers’ production preferences might not mirror listeners’ perception. In fact, hearers may infer the opposite, namely that speakers who assert rather than presuppose p (and refrain from sneaking p into the common ground) signal that p holds up to scrutiny in the subsequent discussion. In this way, hearers may infer that the bare assertion conveys higher degrees of belief than know.

7 Note, we are assuming a relative lack of addressability without going into detail about the role that at-issueness potentially plays here. As was illustrated in the introduction, at-issueness was found to influence the extent to which content is addressable. Similarly, based on Chapter 2, objecting to presupposed at-issue content seemed not more effortful than objecting to asserted at-issue content. However, (i) running one experiment is not enough to make a claim that presupposed content and asserted content are similarly easily addressable when being at-issue, (ii) we are now testing the actual production of presupposition triggers and will keep working with the hypothesis that speakers may exploit the tendency that presupposed content may be more effortful to address than asserted content. In our experiment the content that is either presupposed or asserted will be at-issue with regards to the QUD.
We will call the view that *know* may have a strategic advantage in putting forward propositions the Presupposition Strength Account. In support of this, Lorson et al. (2019) (Chapter 2) provide evidence suggesting that interlocutors are marginally more likely to challenge false information when it was asserted such as in (52b) than when it was presupposed such as in (53b), investigating a variety of presupposition triggers (*stop, regret, to be happy, return, know, discover, to be annoyed*).

(52)  
\begin{align*}
  \text{a. Participant: Have you held any other positions?} \\
  \text{b. Suspect: I used to work for the national gallery in Russia until 2017.}
\end{align*}

(53)  
\begin{align*}
  \text{a. Participant: Have you held any other positions?} \\
  \text{b. Suspect: I stopped working for the national gallery in Russia in 2017.}
\end{align*}

In Lorson et al. (2021)’s (Chapter 3) study, participants were asked to take over the role of a detective who interrogates a suspect (a trained confederate). The participants’ task was to ask the suspect a list of prepared interrogation questions and to compare the suspect’s response with pieces of evidence that were made available to them. Participants were instructed to either challenge or accept the suspect’s answers as appropriate. For the critical items, the suspect would either presuppose (53b) or assert (52b) false information (here, *the suspect worked for the national gallery in Russia until 2017*). The crucial piece of evidence (here, the personal file) would have shown that the suspect worked in Shenzhen, China instead.

The Presupposition Strength Account may explain under which circumstances *know* may be preferred over the bare assertion, namely when speakers do not want p to be discussed further. Such an account would not necessarily make predictions about whether *know* or the bare assertion is epistemically stronger. However, it allows the formulation of a prediction about the variability of knowledge states underlying *know* in non-cooperative contexts. The Presupposition Strength Account predicts that speakers would use *know* for high degrees of belief but sometimes even for low degrees of belief for discourse structuring purposes. To be precise, this account makes a prediction about what the speaker tries to do. However, it is up to the hearer whether they accept presupposed or asserted content which depends on their epistemic vigilance (Sperber et al., 2010) and their presumptions of cooperativity.

In pragmatics it is usually assumed that speakers are cooperative and under this assumption the emergence of additional contextually-determined meanings is explained. The question naturally arises of how pragmatics is supposed to work in an ‘uncooperative’ context.\(^8\) At the level of assertion, it is clear that interlocutors can understand people perfectly well without making any assumptions

\(^8\)We thank an anonymous reviewer for pointing this out.
about their accuracy or trustworthiness as interlocutors – that is to say, interlocutors can recognise an assertion as a falsehood. 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. This kind of speaker behaviour was already acknowledged by Grice (1989): speakers may violate cooperative principles or maxims, for instance, they may violate the maxim of quantity by being uninformative or violate the maxim of quality by lying. 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 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.

4.5 The present study

We have introduced a number of different approaches to capture the production of know and the bare assertion. However, more empirical evidence is needed to understand potential differences between these formulations better. We conducted two experiments to test the competing predictions of the four accounts introduced above. With respect to the three epistemic accounts, we aimed to investigate whether epistemic strength, and the extent to which speakers can use a formulation more or less flexibly, influences the production and comprehension of know and the bare assertion. Regarding the Presupposition Strength account, we aimed to explore whether speakers adapt their communicative strategies depending on the scenario they are in and whether hearers are sensitive to any scenario-driven adjustments by speakers. For this purpose, we conducted a production experiment where participants chose between believe p, know p and p and estimated their confidence in p in two contrasting scenarios. Similarly, in our comprehension study, we asked participants to rate a speaker’s certainty when saying either believe p, know p or p also in two contrasting scenarios.

In order to test the predictions of the Presupposition Strength Account we created a cooperative and an uncooperative scenario, i.e. a scenario where the speaker should be concerned with discussing their knowledge state honestly to reach a joint communicative goal with the hearer versus a scenario where the speaker is not concerned with giving the hearer a maximally honest and objective impression of the speaker’s knowledge state. In order to create a coherent story where it is plausible for a speaker to engage in cooperative and uncooperative communication we used the detective cover story from previous experiments (Lorson et al., 2021) (Chapter 3) with two different scenarios: a briefing with a colleague (cooperative scenario), and an interrogation of a suspect (uncooperative scenario).
We understand a briefing with a colleague to be a semi-standard scenario where interlocutors jointly work towards a communicative goal. We understand the 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. Likewise, the suspect most likely does not want to disclose all information they have 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 are expected to use interrogation strategies and counter-interrogation strategies such as bluffing by overstating their certainty to reach/hide the truth (Granhag & Hartwig, 2008; Luke, 2021).

In the production study, based on Lorson et al. (2021) (Chapter 3), participants are shown evidence that bears upon the likely truth 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 and avoided using quantitative prompts which involve mathematical/statistical/logical reasoning in order to arrive at an event probability estimate, since participants may struggle to understand probabilistically presented information (Kahneman, 2011). We assess 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 study, participants were 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: https://osf.io/3eq9z/.

4.5.1 Experiment 1 [4]: 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 (i) replicate Lorson et al. (2021)’s findings (Chapter 3) that the usage of know depended on context and that know was uttered for higher degrees of belief than believe; and (ii) extend those findings by testing for a difference between presupposing content (i.e. choosing know) and asserting content (i.e. choosing the bare assertion). The experiment and analysis, specified below, were pre-registered; see https://osf.io/3eq9z/.

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 briefed a colleague and interrogated a suspect, relying on pieces of evidence about the suspect’s whereabouts. In this way we elicited expressions that correspond to degrees of belief and we manipulated the context by varying the addressee. 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 produce utterances in two scenarios that differ in their assumed cooperativity (briefing vs interrogation) concerning propositions about which they have a range of confidence levels (low to high).

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.

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 × strength). However, for the analysis, we did not rely on this categorisation of evidential strength, but instead on the participants’ evaluation of that in a 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 (4.1-4.2) and Appendix 7.4.4 for full details.9

9To 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?

Financially the suspect was doing alright.

Evidence:

Summary of Accounts

<table>
<thead>
<tr>
<th>Account Type</th>
<th>Account Number</th>
<th>Ending Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2X000000000001</td>
<td>-13,000</td>
<td></td>
</tr>
</tbody>
</table>

Primary Checking

<table>
<thead>
<tr>
<th>Account Summary</th>
<th>Date</th>
<th>Description</th>
<th>Amount</th>
<th>Ending Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10/23/2019</td>
<td></td>
<td>F 100</td>
<td>€ 12,000</td>
</tr>
</tbody>
</table>

Figure 4.1: Briefing item for the production experiment [experiment 4] (suspect Emily Brown, strong evidence).
The question in the heading of the picture was meant to be a question already asked by the participant. 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 filling the gap in a sentence by choosing between know, believe and the bare assertion highlighted in the briefing (54) and interrogation (55) items. 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.

(54) Briefing item:
   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.

(55) Interrogation item:
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 utterance choice and to create a more natural experience. The response other made up 13.5% of the data.\(^\text{10}\)

The manipulated pieces of evidence ranged from pictures to 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 (54) and (55) a bank statement was provided as strong evidence (Figure 4.1) and a statement of a friend who mentions potential financial difficulties as weak evidence (Figure 4.2); see Appendix 7.4.4 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 4.3) and the participant had to choose the correct answer (here 11am), see (56). The option other was also available.

\(^{10}\)We 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 overall satisfied with the utterance options. When participants chose other, in the briefing, they mainly choose to produce utterances containing alternative uncertainty expressions other than believe such as Johann Smith may have entered via the air ventilation ducts. or referred to the evidence, e.g. Johann 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?.
After the main task, participants 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).

**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 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). We instructed participants to converse with the colleague/suspect about different topics including questions about the case by choosing between utterances 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 then 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. footage of the suspect being at the party). The evaluation task was followed by the voluntary demographic questionnaire.

**Predictions**

We base our predictions about *believe* on Lorson et al. (2021)’s findings (Chapter 3). Their findings suggest that speakers’ choice to produce *believe* is based on their degrees of belief, and furthermore, on the scenario they are in. Thus, we predict that participants will choose *believe* for lower evidentiality ratings than *know*. Furthermore, we predict that participants are more likely to choose *believe* in a cooperative setting (briefing) than in an uncooperative setting (interrogation). We test predictions about the use of the bare assertion and *know* that are based on the accounts of epistemic strength introduced in section 1. We also test predictions about the effect of scenario based on the Presupposition Strength Account introduced in section 2. Note that we do not directly test the differences between high-epistemic versus low-epistemic standards contexts discussed by Epistemic Contextualism and the Knowledge in Context account; we will briefly return to this issue in our discussion for potential future studies. We test the following predictions:

**Knowledge Norm of Assertion**  Speakers should assert p only when they are absolutely certain that p is true. This means, averaged across scenarios, participants are predicted to utter the bare assertion only for evidentiality ratings very close to or equal to 100.

**Epistemic Contextualism**  Speakers may utter *know* as long as the epistemic requirements of the context are met, which may be less than absolute certainty, meaning that *know* can be uttered for a variety of contexts and for varying degrees of belief. Thus, averaged over scenarios, participants are predicted to utter *know* for high evidentiality ratings but not necessarily 100. In comparison with the bare assertion, participants are predicted to utter *know* for lower evidentiality ratings with a wider range than the bare assertion.

**Knowledge in Context Account**  Speakers may utter *know* or the bare assertion as long as the epistemic requirements of the context are met, which may be less than absolute certainty. *Know* is seen as epistemically stronger than the bare assertion. Hence, averaged over scenarios, participants are predicted to utter *know* and the bare assertion for high evidentiality ratings close to but not necessarily equal to 100. However, the mean evidentiality ratings for *know* should be higher than
the mean evidentiality ratings for the bare assertion, assuming that speakers are going to produce the strongest thing that is licensed by the context.

**Presupposition Strength Account: Briefing**  In cooperative settings, speakers should avoid presupposing potentially controversial information and instead assert it given the relative lack of addressability. *Know* should only be uttered if the material is most likely uncontroversial. Thus, participants in the briefing should favour the bare assertion over *know* to make most content available for discussion. This account therefore predicts that *know* will be associated with higher evidentiality ratings on average than the bare assertion, as it is restricted to cases where the proposition is uncontroversial.

**Presupposition Strength Account: Interrogation**  In uncooperative settings, speakers should make use of the discourse structuring properties of *know*, i.e. to signal that they do not want to discuss whether the presupposed content is true or false. Instead, it is signalled that the presupposed content should be accepted as a given and as being already part of the common ground. Furthermore, speakers may utter *know* for low and high degrees of belief for the purpose of employing the discourse structuring properties of *know*. Thus, in the interrogation scenario, participants are predicted to utter *know* more frequently than in the briefing scenario and to favour *know* over the bare assertion. Furthermore, *know* is predicted to occur in conjunction with lower evidentiality ratings in the interrogation than in the briefing, since whether material is controversial or not should matter to a lesser extent.

**Analysis**

The data was analysed 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) and the continuous variable *evidentiality* 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 as interrogation and 1 as 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’ utterance 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 by-expression intercepts were normal distribu-
tions with mean 0 and standard deviation 10. This means that we could be 68% certain that the by-expression 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 et al., 2016). We ran a second model with uninformative priors that yielded similar posterior distributions as the model with weakly regularising priors, see Appendix 7.4.1 for the output.  

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 whereby the first 1000 iterations were disregarded as part of the warm-up phase leading to 12000 iterations available for analysis.

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 et al., 2012).

Results

All participants were above 97% accurate for the control items, which suggests that they paid attention during the experiment. Overall, participants made use of all options, see Table 4.1 for details. The sentential verb believe was used most often followed by the bare assertion and know.

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 and medians of the evidentiality measures for the scenarios briefing and interrogation where almost identical (mean=75/74, sd=24.6/25.7, median=80/80.5 respectively) which was desired. A difference in evidentiality ratings between scenarios would have introduced a confounding variable.

---

11 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 7.4.3.
Table 4.1: Raw counts for each utterance of the production experiment [4] ordered from most frequent to least frequent.

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 rating were 86 and 79.5 respectively), whereas they used *believe* for weaker evidence (mean evidentiality rating was 66.5); see Table 4.2 for details.

Table 4.2: Mean evidentiality ratings by utterance and scenario for the production experiment [4].

Contrasting the by-expression evidentiality ratings for both scenarios reveals that, overall, the median evidentiality ratings across expressions were lower for events mentioned in the interrogation than in the briefing; see Figure 4.4 and Table 4.2.

These observations are supported by the outcome of the analysis. For both *believe* ($\hat{\beta} =-1.51$, CrI:[-1.85, -1.19]) and the bare assertion ($\hat{\beta} =-0.36$, CrI:[-0.67, -0.04]) an increase in evidentiality meant a decrease in log-odds of production; see Table 4.3. 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, *believe* ($\hat{\beta} =0.59$, CrI:[0.39, 0.81]) and the bare assertion ($\hat{\beta} = 0.27$, CrI:[0.05, 0.48]) were less likely chosen over *know* in the interrogation and more likely chosen over *know* in the briefing scenario. At a very high level of certainty (evidentiality = 99.5, which is 1 standard deviation above the mean), speakers are predicted to choose *know* over *believe*
with a probability of 0.44 in the briefing, and 0.72 in the interrogation. Likewise, speakers are predicted to choose *know* over the bare assertion with a probability of 0.46 in the briefing and 0.60 in the interrogation.\textsuperscript{12} Overall, we found the effect of scenario to be less pronounced than the effect of evidentiality.\textsuperscript{13}

\textsuperscript{12}Probability 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.

\textsuperscript{13}This is based on comparing the expected log predictive density of the full model with models that only contained either the scenario or the evidentiality predictor via PSIS-LOO approximation (Pareto smoothed importance sampling leave-one-out approximation; Vehtari and Gelman (2015); Vehtari, Gelman and Gabry (2017)). The approximation disclosed that the full model had the highest predictive accuracy, followed by the model that only included evidentiality as a predictor. The model that only included scenario as a predictor had the lowest predictive accuracy. The table below shows the difference in the models’ expected predictive accuracy (elpd\_diff) and the standard error of the difference in elpd (se\_diff) whereby the preferred model is listed first.

<table>
<thead>
<tr>
<th></th>
<th>elpd_diff</th>
<th>se_diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>full model</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>model with evidentiality</td>
<td>-39.8</td>
<td>9.5</td>
</tr>
<tr>
<td>model with scenario</td>
<td>-132.0</td>
<td>16.1</td>
</tr>
</tbody>
</table>
Table 4.3: Production experiment [4]: 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). \( \hat{\mathcal{R}} \) is a convergence diagnostic which compares the between- and within-chain estimates. Values larger than 1 suggest that the chains have not mixed well.

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Coefficient</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>( \hat{\mathcal{R}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>believe</td>
<td>Intercept</td>
<td>1.18</td>
<td>0.21</td>
<td>0.77</td>
<td>1.60</td>
<td>1.00</td>
</tr>
<tr>
<td>bare assertion</td>
<td>Intercept</td>
<td>0.26</td>
<td>0.16</td>
<td>-0.07</td>
<td>0.57</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Evidentiality</td>
<td>-1.51</td>
<td>0.17</td>
<td>-1.85</td>
<td>-1.19</td>
<td>1.00</td>
</tr>
<tr>
<td>bare assertion</td>
<td>Evidentiality</td>
<td>-0.36</td>
<td>0.16</td>
<td>-0.67</td>
<td>-0.04</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Scenario</td>
<td>0.59</td>
<td>0.11</td>
<td>0.39</td>
<td>0.81</td>
<td>1.00</td>
</tr>
<tr>
<td>bare assertion</td>
<td>Scenario</td>
<td>0.27</td>
<td>0.11</td>
<td>0.05</td>
<td>0.48</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Figure 4.5: Predictions of the model given the data of the production experiment [experiment 4]. The figure contrasts the two scenarios for each utterances. 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 given the data in Figures 4.5 and 4.6. In order to facilitate understanding we back-transformed the 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 the faded area depicts the 95% credible interval and the darker area the 80% credible interval of the fixed effects.

Figure 4.5 shows the predicted model probabilities of using a specific expression 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 between scenarios for the bare assertion. In contrast, the production of *believe* and *know* is predicted to differ 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 4.5](image)

Figure 4.6: Predictions of the model given the data of the production experiment [experiment 4]. The figure contrasts the three utterance 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 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.

Figure 4.6 illustrates the predicted model outcome in different facets than Figure 4.5 to emphasise the predicted differences between utterances – and specifically between *know* and the bare assertion –
in the two scenarios. In the briefing, the model given our data 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 (although the difference between utterances appears less pronounced): the bare assertion is chosen more likely than know. However, Figure 4.6 suggests that when speakers get more confident – at mean evidentiality ratings (74.55) and above – know is predicted to be 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. Based on a reviewers suggestion, we investigated whether the production of know is impacted by an interaction of evidentiality ratings and scenario we carried out a post hoc analysis where we included the interaction (scenario * evidentiality). We had specified and pre-registered a model with main effects without interaction because what was of interest was whether there was an effect of evidentiality (different forms for different evidentiality rates) and an effect of context (different forms for interrogation vs briefing). None of the theoretical claims we are working with make predictions for an interaction (that different evidentiality ratings have a greater impact on form selection in one context than another). For this post-hoc interaction analysis we specified the same priors as for the main effects model.

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Coefficient</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>1-95% CrI</th>
<th>u-95% CrI</th>
<th>( \hat{R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>believe</td>
<td>Intercept</td>
<td>1.25</td>
<td>0.22</td>
<td>0.82</td>
<td>1.68</td>
<td>1.00</td>
</tr>
<tr>
<td>bare assertion</td>
<td>Intercept</td>
<td>0.26</td>
<td>0.17</td>
<td>-0.09</td>
<td>0.58</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Evidentiality</td>
<td>-1.58</td>
<td>0.18</td>
<td>-1.96</td>
<td>-1.24</td>
<td>1.00</td>
</tr>
<tr>
<td>bare assertion</td>
<td>Evidentiality</td>
<td>-0.37</td>
<td>0.17</td>
<td>-0.69</td>
<td>-0.04</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Scenario</td>
<td>0.65</td>
<td>0.12</td>
<td>0.42</td>
<td>0.90</td>
<td>1.00</td>
</tr>
<tr>
<td>bare assertion</td>
<td>Scenario</td>
<td>0.28</td>
<td>0.12</td>
<td>0.04</td>
<td>0.52</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Evidentiality * Scenario</td>
<td>-0.21</td>
<td>0.15</td>
<td>-0.50</td>
<td>0.07</td>
<td>1.00</td>
</tr>
<tr>
<td>bare assertion</td>
<td>Evidentiality * Scenario</td>
<td>0.02</td>
<td>0.14</td>
<td>-0.25</td>
<td>0.31</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 4.4: Production experiment [4]: 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. 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.

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 means that we cannot exclude the possibility that there is no interaction effect; see Table 4.4 and Figure 7.9 in Appendix 7.4.3. However, in order to understand better how all three formulations are affected by the predict-
ors 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 4.7 displays only three reference points of the continuous measure evidentiality. We used the same method as Sóskuthy and Roettger (2020), and an interested reader may look at our code here https://osf.io/3eq9z/.

![Density Plot](image)

Figure 4.7: The density plot for the production experiment [experiment 4] shows the difference in probability of saying one of the formulations in the interrogation relative to the briefing for low (green), average (pink) and high evidentiality ratings (blue).
Figure 4.7 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. However, all three density estimates include 0 which means that we do not have conclusive evidence for an effect of scenario on the bare assertion. In contrast, the production of know seems to be influenced by the factor scenario most clearly for average and high evidentiality ratings opposed to low evidentiality ratings. Similarly, the effect of scenario on the production of believe is more pronounced for average and high evidentiality ratings.

We compared the expected log predictive density of the main effects model with the interaction model via PSIS-LOO approximation (Pareto smoothed importance sampling leave-one-out approximation: Vehtari and Gelman (2015); Vehtari et al. (2017)). The results suggest that both models perform similarly regarding their predictive strength, with the main effects model performing minimally better. While this part of the analysis helps to better understand the interaction of evidentiality and scenario, the introduced results are part of a post-hoc test. Taking this together with the fact that we did not find reliable interaction effects means that further experimental investigations are needed to understand better in what way evidentiality and scenario influence the production of believe/bare assertion/ know.

<table>
<thead>
<tr>
<th></th>
<th>elpd_diff</th>
<th>se_diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>main effects model</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>interaction model</td>
<td>-2.5</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Table 4.5: Production experiment [4]: The table shows the difference in the models’ expected predictive accuracy (elpd_diff) and the standard error of the difference in elpd (se_diff). The preferred model is listed first.

**Interim Discussion**

Focusing first on our results regarding believe, we were able to replicate Lorson et al. (2021)’s findings (Chapter 3). More specifically, we found that participants tended to choose believe for lower evidentiality ratings than know. Additionally, participants chose believe more readily in the briefing than in the interrogation.

**Knowledge Norm of Assertion** The predictions could not be confirmed. According to our results, participants chose the bare assertion on average for lower evidentiality ratings than predicted by this account and for a wide range of evidentiality measures. More precisely, mean evidentiality ratings for the bare assertion were 79.5.
Epistemic Contextualism Our results align with the predictions: *know* was chosen for a wide range of evidentiality measures with a mean of 86.

Knowledge in Context Account Our results support these predictions. Both the bare assertion and *know* were chosen for a variety of evidentiality ratings. In accordance with the predictions, *know* was on average uttered for higher evidentiality ratings than the bare assertion (86 > 79.5).

Presupposition Strength Account: Briefing According to our study, participants indeed favoured the bare assertion over *know* in the briefing. Moreover, in agreement with the account’s predictions, higher evidentiality ratings were needed for a speaker to utter *know* instead of the bare assertion. This seems to suggest that speakers preferred to assert content when it was controversial opposed to presupposing it.

Presupposition Strength Account: Interrogation The predictions could partly be confirmed: *know* was more likely to be chosen in the interrogation than in the briefing. The bare assertion was still uttered more often overall (absolute count = 225) than *know* (absolute count = 205). However, there was a weak tendency that *know* was more likely to be uttered than the bare assertion for higher evidentiality ratings (> 74.55). In agreement with the predictions, *know* was uttered in conjunction with lower evidentiality ratings in the interrogation than in the briefing.

4.5.2 Experiment 2 [5]: Comprehension task

The results of experiment 1 [4] suggest that the production of *know* differs between scenarios, which is compatible with a view on which speakers use *know* strategically. This raises the question of whether hearers take the speakers’ strategies into account while forming an understanding of the speakers’ epistemic state. In experiment 2 [5], we test this by eliciting hearers’ assessments of speaker certainty in the scenarios used in experiment 1 [4], to explore whether these are based solely on the speakers’ utterance choice (bare assertion vs. *believe* vs. *know*) or also on the communicative setting. The experiment and analysis, specified below, were pre-registered; see https://osf.io/3eq9z/.

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.
Participants

Participants (N=121) were recruited from Prolific following the approach in experiment 1 [4]. 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 person preferred not to disclose this.

Design and Materials

Design and materials were used from experiment 1 [4] and were only minimally changed. 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 either one of the utterance alternatives yielding a 2×3 design (scenario × utterance). As in the production task, we used pictures to set the scene for 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 Appendix 7.5.3 for full details. The pictures were assembled like a comic strip. In the briefing (57), Colleague1 answers an already asked question (speech bubble in first picture), and then Colleague2 responds (speech bubble in second picture). In the interrogation (58), the first picture depicts the suspect’s answer in form of a speech bubble and the second picture depicts the colleague’s statement. The utterance of the second turn was the critical utterance which was either the bare assertion (57a) or contained believe (57b) or know (57c).

As in experiment 1 [4], 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 item, 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 participants were asked to assess the speaker’s certainty by adjusting a slider ranging from 0 (not at all certain) to 100 (very certain).

(57) 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.

(58) 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 an utterance. Again, we included 10 attention checks; see (56) for more details.

**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 4.5.2. 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 (filler items) and that their task as a detective in training was to observe and assess their colleagues (critical items). The experimental task was followed by the voluntary, demographic questionnaire. The experiment lasted approximately 20 minutes.

**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 are behaving
in accordance with their belief that speakers are acting in particular ways. For instance, if hearers assume that speakers are bound by the Knowledge of Assertion, they should assign certainty = 100 to bare assertions. Thus, we will derive our predictions from the previously discussed accounts which are oriented towards production and assume that hearers assess speaker meaning while being aware of the communicative choices and strategies that are available to speakers.

We base our predictions for assessed speaker certainty for believe on the results that were obtained here and in Lorson et al. (2021) (Chapter 3): hearers are predicted to assign lower certainty to a speaker uttering believe than to a speaker uttering know. However, the difference between know and the bare assertion is again more controversial. We test the following predictions:

**Knowledge Norm of Assertion** Speakers assert p only when they are absolutely certain that p is true. This means participants are predicted to assign certainty values very close to or equal to 100 to a speaker uttering the bare assertion.

**Epistemic Contextualism** Speakers may utter know as long as the epistemic requirements of the context are met, which may be less than absolute certainty. Thus, averaged over scenarios, participants are predicted to assign high certainty values close to but not necessarily equal to 100 to a speaker uttering know. In comparison with the bare assertion, participants are predicted to assign lower certainty values with a wider range to a speaker uttering know than a speaker uttering the bare assertion.

**Knowledge in Context Account** Speakers may utter know or the bare assertion as long as the epistemic requirements of the context are met which may be less than absolute certainty. Know is seen as epistemically stronger than the bare assertion. Hence, averaged over scenarios, participants are predicted to assign high certainty values close to but not necessarily equal to 1 to a speaker uttering know or the bare assertion. However, mean speaker certainty ratings for know should be higher than the mean speaker certainty ratings for the bare assertion.

**Presupposition Strength Account: Briefing** In cooperative scenarios, speakers should avoid presupposing potentially controversial information and instead assert it given the relative lack of addressability. Know should only be uttered if the material is most likely uncontroversial. Thus, participants in the briefing are predicted to assign higher confidence ratings to a speaker uttering know (values close/equal to 1) than a speaker uttering the bare assertion, assuming hearers are aware that speakers would only resort back to presupposing information in the case where it is uncontroversial and the speaker is sure of that.
Presupposition Strength Account: Interrogation  In the uncooperative scenario, speakers may make use of the discourse structuring properties of know. Thus, in the interrogation scenario, participants are predicted to assign on average lower certainty values to a speaker uttering know than a speaker uttering the bare assertion, assuming hearers are aware of the speakers’ strategies in the interrogation. Furthermore, participants are predicted to assign lower certainty values to a speaker uttering know in the interrogation than in the briefing.

Analysis
The data was analysed 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, $S_{\text{certainty}}$ is the dependent variable speaker certainty and $n$ is the number of observations (here 2420).\textsuperscript{14}

$$[(S_{\text{certainty}}/100) \times (n - 1) + 1/2]/n$$

The experimental factor scenario (briefing/interrogation) and utterance (believe/know/bare assertion), were included to predict the assessed speaker certainty. The factor scenario was sum-coded: -1 as interrogation and 1 as briefing. For the factor utterance the following sum coding was used:

<table>
<thead>
<tr>
<th>Utterance I</th>
<th>Utterance II</th>
</tr>
</thead>
<tbody>
<tr>
<td>believe</td>
<td>1</td>
</tr>
<tr>
<td>assertion</td>
<td>0</td>
</tr>
<tr>
<td>know</td>
<td>-1</td>
</tr>
</tbody>
</table>

Table 4.6: Comprehension experiment [5]: Sum coding for the predictor utterance

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

\textsuperscript{14}We decided to do 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.
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 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 7.5.1 for the output.\footnote{15}{See footnote 11.}

We used the NUTS sampler (Hoffman & Gelman, 2013) as before, with four sampling chains, each collecting 4000 iterations whereby the first 1000 iterations were disregarded as part of the warm-up phase leading to 12000 iterations available for analysis. We will report the posterior mean $\hat{\beta}$ and the 95% credible interval (95%-CrI).

**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 did not vary in their evidential strength between scenarios (mean=0.69/0.7, sd=0.25/0.24, median=0.75/0.75 respectively).

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

Table 4.7: Comprehension experiment [5]: Mean speaker confidence ratings by utterance

Furthermore, speaker confidence ratings varied between utterances. 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 4.7 for details.

Figure 4.8: By-expression speaker confidence ratings of the comprehension experiment [experiment 5] for each scenario (briefing in blue, interrogation in red). The figure shows the median of the speaker confidence ratings (line) and the upper quartile and lower quartile (box). Whiskers extend to the scores outside the the quartiles. Dots represent outliers.

Contrasting the by-utterance confidence ratings for both scenarios reveals only a small difference between the scenarios. A speaker uttering believe received on average higher confidence ratings in the interrogation than in the briefing. Median confidence ratings for speakers uttering know and the bare assertion are similar across scenarios; see Figure 4.8.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>posterior mean</th>
<th>Standard Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>Ṙ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (grand mean)</td>
<td>1.05</td>
<td>0.09</td>
<td>0.86</td>
<td>1.24</td>
<td>1.00</td>
</tr>
<tr>
<td>Utterance I</td>
<td>-0.82</td>
<td>0.08</td>
<td>-0.97</td>
<td>-0.68</td>
<td>1.00</td>
</tr>
<tr>
<td>Utterance II</td>
<td>0.16</td>
<td>0.06</td>
<td>0.05</td>
<td>0.28</td>
<td>1.00</td>
</tr>
<tr>
<td>Scenario1</td>
<td>-0.07</td>
<td>0.04</td>
<td>-0.16</td>
<td>0.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 4.8: Comprehension experiment [5]: 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 (4.6). 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. Utterance1/UtteranceII = -1: $-0.82 \times (-1) + 0.16 \times (-1) = 0.66$. The effect scenario is the change in log-odds for the briefing (-1 interrogation, 1 briefing).
These observations are supported by the outcome of the analysis. Averaged over scenarios, participants assigned certainty values below average (grand mean = 0.74 on the original scale) when assessing a speaker uttering *believe* \( (\hat{\beta} = -0.82, \text{CrI:}[-0.97, -0.68]) \) and slightly above average when assessing a speaker uttering the bare assertion \( (\hat{\beta} = 0.16, \text{CrI:}[0.05, 0.28]) \), see Table 7.6. Participants assigned the highest certainty values to a speaker uttering *know* (on average 0.66 above the grand mean/intercept on the log-odds scale).

![Predicted speaker certainty by utterance and scenario](image)

Figure 4.9: Model predictions given the data of the comprehension experiment [experiment 5]. Log-odds were back-transformed to speaker certainty (y-axis). This figure shows the assigned speaker certainties by utterance and scenario (briefing in blue, interrogation in red) with error bars, representing the 95% credible intervals.

Regarding effects of scenario, there was only a weak tendency for participants to assign higher certainty values to a speaker in the interrogation than in the briefing \( (\hat{\beta} = -0.07, \text{CrI:}[-0.16, 0.01]) \), see also Figure 4.9. These results have to be interpreted with care, since zero lies inside the parameter’s estimated credible interval.

As before we will run a post hoc analysis where we include the interaction (scenario \( \times \) utterance) in order to understand how the assigned speaker certainties vary between scenarios for the three formulations separately. Again, we had specified and pre-registered a model with main effects without interaction based on the predictions that utterance and scenario would have an additive effect on the assigned speaker certainty. For this post hoc analysis we specified the same priors as for our main effects comprehension model. The results suggest that there is no evidence for an
interaction of believe ($\hat{\beta} = -0.03, \text{CrI}:[-0.09, 0.04]$) or the bare assertion ($\hat{\beta} = -0.00, \text{CrI}:[-0.07, 0.06]$) with scenario in predicting speaker certainty.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>posterior mean</th>
<th>Standard Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>(\hat{R})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (grand mean)</td>
<td>1.06</td>
<td>0.09</td>
<td>0.87</td>
<td>1.24</td>
<td>1.00</td>
</tr>
<tr>
<td>Utterance I</td>
<td>-0.83</td>
<td>0.08</td>
<td>-0.98</td>
<td>-0.68</td>
<td>1.00</td>
</tr>
<tr>
<td>Utterance II</td>
<td>0.16</td>
<td>0.06</td>
<td>0.05</td>
<td>0.27</td>
<td>1.00</td>
</tr>
<tr>
<td>Scenario1</td>
<td>-0.07</td>
<td>0.04</td>
<td>-0.16</td>
<td>0.01</td>
<td>1.00</td>
</tr>
<tr>
<td>Utterance I * Scenario1</td>
<td>-0.03</td>
<td>0.03</td>
<td>-0.09</td>
<td>0.04</td>
<td>1.00</td>
</tr>
<tr>
<td>Utterance II * Scenario1</td>
<td>-0.00</td>
<td>0.03</td>
<td>-0.07</td>
<td>0.06</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 4.9: Comprehension experiment [5]: 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 (4.6).

We used the same method as before to illustrate how the effect of scenario on the assessed speaker confidence varies between the three utterances. Figure 4.10 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. If there was an effect it would most likely be driven by believe.
Figure 4.10: Comprehension experiment [experiment 5]: 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).

We compared the expected log predictive density of the main effects model with the interaction model via PSIS-LOO approximation. The results suggest that both models perform fairly similarly regarding their predictive strength with the model without interaction performing slightly better.

<table>
<thead>
<tr>
<th>Model</th>
<th>elpd_diff</th>
<th>se_diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>main effects model (comprehension)</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>interaction model (comprehension)</td>
<td>-2.8</td>
<td>6.7</td>
</tr>
</tbody>
</table>

Table 4.10: Comprehension experiment [5]: The table shows the difference in the models’ expected predictive accuracy (elpd_diff) and the standard error of the difference in elpd (se_diff). The preferred model is listed first.

Interim Discussion

In agreement with the production data of Lorson et al. (2021) (Chapter 3), we found that participants assigned lower certainty to a speaker uttering believe than to a speaker uttering know. In what follows we will discuss our results for the difference in assigned speaker certainty between know and the bare assertion.
Knowledge Norm ofAssertion In contrast to the predictions, we found that on average participants assigned certainty values of 72.9 to a speaker uttering the bare assertion.

Epistemic Contextualism Our results are in agreement with this prediction, since participants assigned a wide range of certainty values with a mean of 80.3 to speakers uttering know.

Knowledge in Context Account Our results align with the predictions: participants assigned high certainty values to speakers uttering the bare assertion or know and not only for maximal certainty (100). Moreover, on average participants assigned higher certainty to a speaker uttering know than the bare assertion.

Presupposition Strength Account: Briefing In alignment with the predictions, in the briefing, participants assigned higher certainty values to a speaker uttering know than to a speaker uttering the bare assertion. This is weak evidence, although it does run in the predicted direction.

Presupposition Strength Account: Interrogation The predictions are not borne out by the results. In the interrogation, participants still assigned higher certainty values to a speaker uttering know than to a speaker uttering the bare assertion. 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 [5] misunderstood the participants/speakers of experiment 1 [4]: Participants/hearers assigned higher speaker confidence values than the speakers’ actual degrees of belief. However, we cannot be sure, since the coefficient lies between -0.16 and 0.01 with 95% probability which includes 0.

4.6 Discussion

4.6.1 Epistemic Accounts

Our results suggest that speakers use know, the bare assertion and believe for content with successively lower evidentiality scores and that 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.

The bare assertion was chosen more often than know in the production study when participants were somewhat uncertain. Similarly, in the comprehension study, participants interpreted know to convey higher speaker certainty than the bare assertion. In this respect, our results align with what
we term the Knowledge in Context Account of DeRose (2002): speakers behave as though know is epistemically stronger than the bare assertion. Furthermore, the wide range of both measures – evidentiality in the production task and speaker confidence in the comprehension task – for both the bare assertion and know suggests that both formulations can be used in a variety of contexts for varying degrees of belief. This goes against the predictions by the Knowledge Norm of Assertion which postulated that the bare assertion is only to be used by speakers who possess maximal certainty.

Furthermore, our production data provide evidence that the production of know and the bare assertion differ between contexts, which aligns well with the predictions of Epistemic Contextualism and the Knowledge in Context Account. For example, speakers are willing to utter know for lower degrees of belief in uncooperative settings than in cooperative settings. However, we did not test the context specific predictions of these accounts. For example, Epistemic Contextualism and the Knowledge in Context Account discussed whether speakers utter know or the bare assertion more readily and for lower degrees of belief in contexts with low epistemic standards opposed to contexts with high epistemic standards. We will illustrate two potential ways of interpreting our data regarding the low- versus high-stakes distinction.

Our results would align with the the predictions of Epistemic Contextualism and the Knowledge in Context Account if we were to assume that the interrogation scenario involves lower stakes than the briefing scenario. This would explain why speakers seem to utter know for lower degrees of belief in the interrogation – the low-stakes setting – than in the briefing – the high-stakes setting. The interrogation could involve lower stakes than the briefing on the social level, since it potentially involves less risk of social sanction if the speaker utters something false. Being in the role of the interrogator makes it acceptable to follow strategies such as bluffing and overstating one’s certainty. In contrast, saying something false in the briefing is less acceptable, since the interlocutors have to maintain mutual trust.

However, one could equally well argue that the briefing involves lower-stakes than the interrogation, since the outcome of a briefing does not impact a criminal case as heavily as the outcome of an interrogation. What has been gathered from a suspect in an interrogation has to prevail in front of a criminal court. If the briefing indeed is the low-stakes context our results would go against Epistemic Contextualism and the Knowledge in Context account. Thus, to shed more light onto this matters future studies should test this further, for instance by contrasting the production of cooperative speakers in low- and high-stakes contexts (e.g. a conversation between friends versus making a statement in court).
4.6.2 Presupposition Strength Account

Our production results about contextual differences fit with the predictions of the Presupposition Strength Account. Speakers seem to utter *know* differently depending on scenario, and differently from the bare assertion. More specifically, speakers uttered *know* for higher degrees of belief in the briefing than in the interrogation where speakers uttered *know* more readily. Furthermore, we found a tendency that speakers preferred *know* over the bare assertion in uncooperative settings for high degrees of belief. For cooperative situations this suggests that speakers avoid presupposing potentially controversial content and tend to presuppose only content that is most likely uncontroversial. For uncooperative situations this may mean that speakers are able to exploit the discourse-structuring properties of the presupposition trigger *know*. By presupposing content, the speaker acts as if that content was already accepted by the interlocutors, and thereby signals that it should not be up for debate. However, as was mentioned before, the observation that *know* was favoured over the bare assertion in the interrogation for high degrees of belief has to be interpreted with caution and needs further testing. We only found weak evidence to support the claim as part of a post-hoc analysis.

Since we did not find that speakers favoured *know* overall over the bare assertion in the interrogation, suggests speakers’ leeway to utter *know* more readily and for potentially controversial content is restricted. This may make sense if one considers that employing strategies such as bluffing successfully requires the speakers to avoid be questioned and to come across as reliable. Thus, an account that captures the production of *know* and the bare assertion should take into account their epistemic strength.

Whereas speakers seem to exploit the discourse structuring properties of the presupposition trigger *know*, we did not find evidence that hearers recover these production strategies. The coefficient for scenario lies between -0.16 and 0.01 with 95% probability which includes 0. This can either mean that hearers assign higher speaker confidence in the interrogation opposed to the briefing or that hearers do not distinguish between scenarios when assessing speaker confidence. Both outcomes would suggest that speakers may employ their strategies successfully in uncooperative interactions. This interpretation would align with Lorson et al. (2019)’s findings (Chapter 2) which indicate that speakers are able to manipulate hearers by using presuppositions strategically. They found that hearers were less likely to object to false information when it was presupposed than when it was asserted.

However, 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 where 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 pretending 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.

Another way of interpreting the results would be to take into account politeness considerations that were briefly mentioned before. Following the argumentation of hedging accounts and politeness theory (e.g., Fraser, 1975; Brown & Levinson, 1987), it could be argued that cooperative speakers tend to downtone their utterances when communicating their certainty in order to be polite. By contrast, in an uncooperative scenario speakers might feel less obliged to engage in facework especially when they are in a high-power position. Thus, instead of hedging their statements, speakers communicate their degrees of belief directly. However, while this analysis explains the use of believe in conjunction with high evidentiality ratings in the briefing scenario, it is not clear that it is an appropriate analysis of the bare assertion, which is not normally considered to be downtoned in any respect. Further work would be required to determine whether discourse participants ever consider it to be a comparatively polite alternative to the corresponding know form.

4.6.3 A revised account

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

First, such a model should capture that believe, the bare assertion and know are produced for content with successively lower evidentiality scores. This may be achieved by assuming a threshold semantics (Lassiter, 2017; Yalcın, 2010): For any expression there exists some threshold \([0, 1)\), and an utterance containing that expression is true if the probability of the event it describes exceeds
this threshold. For example, if the threshold for *believe* is 0.6, then the utterance ‘I believe that the suspect went to the party yesterday’ is available for a speaker who believes that the probability of the suspect having gone to the party yesterday exceeds the threshold of 0.6. Thresholds may vary between speakers but the general order of thresholds should be *believe*<bare assertion<*know* based on what we found about the relative ordering of the formulations’ epistemic strength. Furthermore, since speakers generally aim to be informative, following Grice (1975)’s maxim of quantity, they should choose the utterance with the highest threshold possible: speakers that have degrees of belief that exceed the threshold of *believe* and the bare assertion should choose the bare assertion, since this would be the most informative utterance concerning the speakers’ knowledge state.

Second, both *believe* and *know* can be used strategically by speakers who follow additional communicative goals other than being informative. As was discussed earlier, *believe* (among many other formulations) can be uttered in cooperative contexts to hedge statements in order to be polite. In that sense, speakers’ communicative goal is primarily to be polite and secondarily to be informative. Thus, speakers may choose to utter *believe* (as long as their degrees of belief exceed the threshold for *believe*) despite having degrees of belief that exceed the threshold of the bare assertion because they no longer aim to be maximally informative. Our results suggests that it is unlikely that speakers disregard thresholds completely: *believe* was still highly unlikely chosen for high degrees of belief. Such a model should be extensible to any hedging tools, i.e. *think*, *may*, *probably* etc.

Similarly, as our results suggest, the discourse structuring properties of *know* may be exploited by speakers who aim to divert the subsequent discussion away from the presupposed content by acting as if though it was already shared knowledge. Hence, the overall communicative goal is not to be informative but to deceive. Again, it seems unlikely that speakers disregard thresholds completely because we did not find that *know* was uttered more likely overall in the interrogation. Instead our data seem to align more with the assumptions that speakers whose degrees of belief exceeded the bare assertion threshold but not necessarily the threshold for *know* chose to utter *know* nonetheless.

Considering our data, the hearer point of view may simply be captured by assuming threshold semantics. However, research has suggested that hearers are aware that speakers can hedge their statements (Bonnefon & Villejoubert, 2006). Thus, both the comprehension of *believe* and *know* has to be investigated further to understand in which way hearers’ interpretation is affected by the speakers’ communicative goals.
4.7 Conclusion

In this experiment, we tested the production and comprehension of the formulations *believe/*bare assertion/*know* in two contrasting scenarios. We found that speakers choose between these three formulations based on their degrees of belief, and likewise, that hearers infer the speakers’ confidence based on their utterances. Regarding the formulations’ epistemic requirements and strength, our findings suggest the following ranking: *believe*< bare assertion< *know*. These findings support (DeRose, 2002)’s account but speak against a strict formulation of the Knowledge Norm of Assertion (Williamson, 2000). Furthermore, our findings shed more light on speakers’ potential motivation to presuppose content rather than assert it. Our findings are compatible with a view on which *know* is used strategically by speakers in order to forestall further discussion of the proposition in question. However, whereas speakers and hearers agree in the relative epistemic ordering of the formulations, hearers do not appear consistently to take account of the speakers’ strategies. As mentioned in the discussion, this may be due to the experimental design of the comprehension study where participants were in a bystander role opposed to participants in the production experiment. 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 (i) the formulations’ epistemic strength, and (ii) the strategic properties of *believe* and *know*.

4.8 What’s next?

The production results of experiment [4] suggest that speakers seem to presuppose content more readily – i.e. more frequently and for lower degrees of belief – when speaking to a potentially uncooperative interlocutor who might want to deny that content as opposed to a cooperative interlocutor. In contrast, the production of the bare assertion did not suggest such a pattern. The findings of experiment [4] align with the results of experiments [2–3] where the production of *know* was contrasted with the production of (un)certainty expressions such as *believe*. Taken together, the results of Part II seem compatible with both of the aforementioned interpretations: (i) speakers utter *know* more readily in uncooperative settings and thereby overstate their knowledge, AND (ii) speakers utter *believe* more readily in cooperative settings and thereby understate their knowledge in order to come across as polite. Thus, both expressions may be used strategically with opposite effects.

Part III of this thesis investigates the production of additive presupposition triggers such as *too*, which has been argued (i) to be only felicitous when there is a suitable antecedent in the preceding context, (ii) to be obligatory if there is a suitable antecedent in the preceding context, (iii) to fulfil
discursive functions such as highlighting parallelism between utterance content. The potential obligatory use and discursive functions of additives is investigated by testing the influence of discourse factors such as Similarity and Proximity between antecedent content and the content of the utterance containing the additive on the production of additives. Furthermore, Part III explores whether the speakers’ communicative goals influence whether to produce additives or not. Thus, Part III tests whether the use of additives can be understood as a move to converge (i.e. socially align) with the antecedent speaker by investigating if speakers would omit additives when an additive might signal alignment with an impolite antecedent speaker. The results of Part III shall inform the overall research question of this thesis – i.e. *What motivates speakers to presuppose rather than assert content?* – as well as illustrate potential differences in the way presupposition triggers interact with the surrounding discourse which will be expanded on in the Discussion.
Part III

Discourse Managing
Chapter 5

The discursive function and presuppositional nature of additives

5.1 Abstract

The production of additives such as *too* has been argued to be obligatory if the immediate context contains a suitable antecedent such that the presupposition triggered by additives is satisfied. Similarly, additives have been claimed to fulfil discursive functions such as highlighting parallelism between utterance content. With two production experiments we investigated the potential obligatoryness and discursive functions of additives, by manipulating the antecedents’ salience focusing on the factors Similarity and Turn Distance. We furthermore explored whether the use of additives can be understood as a move to converge (i.e. socially align) with the antecedent speaker by investigating whether speakers would omit additives when an additive might signal alignment with an impolite antecedent speaker. Overall, the results of our two experiments suggest that while the production of additives seems to depend on the antecedent’s salience, additive production was not as frequent as expected if additives were indeed obligatory. More specifically, our findings suggest that (i) in alignment with previous results on similarity, speakers tend to utter additives more frequently when their utterance’s content matches the content of a previously formulated utterance; and (ii) speakers use additives more frequently when the matching utterance directly precedes their utterance. Furthermore, results of experiment II suggest that (iii) speakers deliberately drop the use of additives when wanting to diverge from an impolite speaker. Our findings lend support to models in which speakers use additives as a grounding tool to organise the discourse and maintain discourse coherence.
5.2 Introduction

This paper investigates the production of additive particles such as *too*, which are argued to presuppose a propositional alternative. For example, (59a) as opposed to (59b) is assumed to presuppose that someone other than Donald saw Dune. We will refer to the utterance containing *too* as the host of *too*. If the immediate context entails a suitable antecedent – here, that someone else other than Donald watched Dune – the production of *too* is argued to be obligatory (e.g. Heim, 1991; Percus, 2006). In (59), Speaker1’s question is a suitable antecedent: someone else, namely Lisa, watched Dune. Omitting *too* in (59b) would lead to a marked discourse, indicated by #, whereby (59b) comes across as a stand-alone utterance detached from the previous dialogue or contrastive to the preceding utterance.

(59) Speaker1: Was Lisa the only one who watched Dune?
   a. Speaker2: Donald saw Dune, too.
   b. Speaker2: #Donald saw Dune.

However, the actual production of *too* seems to be more complex. For example, in (60b) the production of *too* seems optional despite there being a plausible antecedent entailed in Speaker1’s utterance (i.e. that Speaker1 has seen Dune).

(60) a. Speaker1: I enjoyed Dune: has anyone else seen it?
   b. Speaker2: I saw it (too) last week.

This study investigates under which circumstances speakers produce additive presupposition triggers such as *too* to understand what contexts (if any) necessitate their presence and how sensitive they are to discourse factors. More specifically, this paper will examine speakers’ production choices to test whether the antecedents’ salience influences the use of additives focusing on (i) Similarity and (ii) intervening Turn Distance between antecedent and host sentence. Furthermore, this paper will explore (iii) whether speakers omit *too* – and as such disengage from signalling discourse relations – to socially diverge from an impolite antecedent speaker and their utterance. We therefore test the factor Politeness, contrasting socially neutral contexts with contexts in which the antecedent speaker is impolite. These three factors were examined by conducting two production experiments; one following an open-choice and one following a forced-choice paradigm.
5.3 The production of additives

5.3.1 Producing additives felicitously

The production of additives is often discussed from two perspectives: (i) the minimum contextual requirements that have to be met for additives to be uttered felicitously, (ii) the contextual requirements that make them presumably obligatory. Considering (i), Kripke (2009) proposed that additives not only presuppose the existence of a propositional alternative but also have an anaphoric component which requires that alternative to be pre-mentioned. Thus, the felicitous use of additives requires a suitable antecedent in the preceding context. What counts as a suitable antecedent has been discussed by different accounts, e.g. accounts such as Beaver and Zeevat (2007) argue that a propositional alternative in the form of individual plus matching property have to be present in the context. For example, for (61), there has to be a salient individual in the preceding context who has the property of being a student.

(61) Lisa is a student, too.

In contrast, Geurts and van der Sandt (2004) argue that additives trigger two presuppositions (62a and 62b, in their example), and that only the first presupposition (62a) is difficult to accommodate. Therefore, the felicitous use of additives is argued to require a salient individual (someone other than the vicar) in the preceding context; however, a salient property (i.e. that x is depressed) in the preceding context is not deemed necessary in the sense that this presupposition can be interpreted via accommodation.¹

(62) The vicar is depressed, too.
   a. There is some person x other than the vicar
   b. x is depressed

The question remains of what counts as a relevant context. Whereas empirical research mostly investigated ‘immediate’ contexts where the antecedent either immediately precedes or is part of the host sentence (e.g. Amsili, Ellsiepen & Winterstein, 2016; Grubic & Wierzba, 2019), Kripke (2009) argues that in order for additives to be used felicitously, the antecedent has merely to be part of the interlocutors’ ‘active context’, which is said to contain the salient material that is on the interlocutors’ minds. Thus, according to Kripke (2009), additives are felicitous even if the antecedent was mentioned a while back or not at all as long as the antecedent is salient. It remains

¹It has to be noted that this is not true for all cases. Geurts and van der Sandt (2004) only claim that whereas (62a) resists accommodation, (62b) can occasionally be accommodated. Thus, in accordance with (Grubic & Wierzba, 2019)’s results, there must at least be a salient individual for too to be used felicitously and for the presupposition to be accommodated.
to be investigated which factors influence the salience of an antecedent and how proximity affects salience, if at all.

5.3.2 Producing additives obligatorily

Considering (ii), the production of additives is said to be obligatory if there is a suitable antecedent in the preceding context (cf. Krifka, 1998; Zeevat, 2003; Sæbø, 2004). For example in (63), Speaker2 should utter the particle too (63a), since the preceding utterance (63) is a suitable antecedent. In contrast, (63b) seems infelicitous to utter.

(63) Speaker1: Lisa finished her PhD in Linguistics.
   a. Speaker2: Mattia finished his PhD in Linguistics, too.
   b. Speaker2: #Mattia finished his PhD in Linguistics.

A mechanism that is thought to explain the obligatory use of additives and other presupposition triggers is Maximize Presupposition (e.g. Heim, 1991; Percus, 2006). According to Maximize Presupposition, a given utterance, e.g. (63b) competes with all its extended versions with presupposition triggers, e.g. (63a). The utterance with the presupposition trigger (63a) will win over versions lacking the presupposition trigger (63b) if the presupposition is satisfied in the context. Not uttering too in a context in which the presupposition of too is satisfied leads to the so-called anti-presupposition effect (Percus, 2006): hearing (63b), interlocutors are argued to infer the anti-presupposition that Mattia is the only salient individual who finished their PhD in linguistics. This anti-presupposition, however, clashes or contrasts with the preceding utterance. In order to avoid such a contrast, too should be uttered, such as in (63).

However, additives have been claimed to differ from triggers such as definite articles in that speakers are not always obliged to insert too even if its presupposition is met in the preceding context. A factor that has been found to influence the use of additives is discourse structure. For example, the use of too is optional in enumerations or lists, see (64).

(64) a. Q: Who in this class is sick today?
   b. A: Lisa is sick. Donald is sick. Mattia is sick. Sarah is sick.

Accounting for cases such as (64), Amsili and Beyssade (2009) suggested a weaker version of Maximize Presupposition that takes into account the influence of discourse structure on additives. In another mechanism that aims to explain the obligatory use of too is the Obligatory Implicatures principle (e.g. Krifka, 1998; Sæbø, 2004; Bade, 2016). The insertion of additives is understood to be needed to cancel exhaustivity implicatures that would clash with the preceding discourse.
their version of Maximize Presupposition, they claim that additives are only obligatory if the utterance with additive presupposition trigger brings strictly more satisfied presuppositions than the utterance without the trigger. An enumeration or list structure is claimed to force the second and following utterance (here, beginning with ‘Donald is sick.’) to be linked to the context which already entails that more than one individual has a given property (for (64b) being sick). Since the utterance without trigger already entails a propositional alternative due to being linked to the context, a sentence with trigger (e.g. ‘Donald is sick, too.’) does not bring more presuppositions than an utterance without trigger. A discourse structure other than lists/enumeration discussed by Amsili and Beyssade (2009) is the CONTRAST/PARALLEL discourse relation, see (65), in which the relation itself entails a similarity between the two propositions.

(65) He was there yesterday, he is there today.

Although Amsili and Beyssade (2009)’s Maximize Presupposition extension seems promising in explaining the occurrences of additives for specific discourse structures, it is indeed limited to those specific structures (lists, enumerations etc.).

5.3.3 The discursive function of too

Signalling similarity

In order to understand under which circumstances speakers utter additives outside of specific discourse structures, recent research has focused on the possible role that additives may play in discourse. So far, a range of different discourse related features have been tested. In a corpus study using the London-Lund Corpus of Spoken English (LLC), Spenader (2002) found that too served mainly the purpose of signalling parallel information. Similarly, Amsili et al. (2016) found that aussi (French too) signals similarity of the sentence containing aussi – the host sentence – and its antecedent. Varying degrees of similarity between antecedent and host were tested based on the assumption that similarity is marked to a greater extent when the host is reduced via ellipsis or anaphora. For example, in one condition the host sentence was identical to the antecedent (see (66a) for the English equivalent), whereas in the more reduced manipulations the objects were replaced gradually by pronouns (66b-66c).

(66) Jean has shown his car to Paul and Lea...
   a. ...has shown her car to Paul, too.
   b. ...has shown her car to him, too.
   c. ...has shown it to him, too.
Amsili et al. (2016) found that the more reduced the host sentence was – and arguably more similar to the antecedent\(^3\) – the more *aussi* was preferred by hearers. From the speakers’ perspective this would mean that with increasing similarity the pressure to utter additive particles such as *too* increases.

(67) John was almost on time.

a. Mary was on time, too.

b. ?Mary was late, too.

This feeds into Winterstein (2010); Winterstein and Zeevat (2012)’s argument that the insertion of *too* depends on argumentative identity. According to their theory, additives can be used felicitously as long as the host and antecedent argue in the same direction. It is assumed that speakers’ utterances are argumentatively oriented towards to some goal the speakers try to defend. Some lexical items are assumed to have inherent argumentative effects, e.g. *almost* combines with a gradable predicate and conveys the negation of its host sentence while retaining some of its argumentative properties, illustrated by Jayez and Tovena (2008). For example, for (67b), it is argued that although *being almost on time* entails *being late*, *being almost on time* retains the argumentative properties of *being on time*. Since (67b) is argumentatively opposed to (67), and assuming that *too* requires argumentative similarity between host and antecedent, (67a) seems more felicitous than (67b).

**Discourse management**

Besides similarity, Eckhardt and Fränkel (2012) argue that additives may be used to manage the discourse. Their account is based on their investigation of *auch* (German *too*) in which they contrasted two different text types: story versus protocol. Eckhardt and Fränkel (2012) asked participants to describe four pictures of a comic strip one after each other. The pictures illustrated two protagonists that would occasionally perform similar actions at different times such as brushing their teeth. Participants were either instructed to describe the comic strip as if they were telling a children’s story or they were instructed to describe what happened at specific times, similar to a protocol. Eckhardt and Fränkel (2012) found that *auch* was used significantly less often in the protocol condition in which the time stamps served as contrastive topics, see (68). Although there is a suitable

---

\(^3\)Amsili et al. (2016) argue that while the full form *Paul* in (66a) may refer to a second individual named Paul (other than the Paul in (66)), the pronominalised form *him* in (66b) refers more clearly to the same individual as Paul in (66). Thus, due to pronominalisation, (66b) is more similar to (66) than (66a). Note that this argumentation is slightly different from Kaplan (1984)’s account, who first observed the increased obligatoriness of additives in anaphorically reduced host sentences. According to Kaplan (1984), reduced host sentences introduce greater prominence of contrastive topic than full host sentences, and thereby, higher pressure to use additives to emphasise the similarity between antecedent and host.
antecedent in the context (Otto is brushing his teeth) which would make *too* obligatory in the last sentence (Fred is brushing his teeth), leaving out *too* doesn’t lead to markedness of the discourse.

(68) 9:00 Fred is knitting. 10:00 Otto is brushing his teeth. 11:00 Otto is reading. 12:00 Fred is brushing his teeth.

Modifying Amsili and Beyssade (2009)’s version of the Maximize Presupposition account, Eckhardt and Fränkel (2012) claim that Maximize Presupposition is a principle to communicate meta-information about the discourse whereby uttering additives signals that the speaker has kept track of the preceding utterances in the discourse. Based on their results they claim that the obligatoriness of additives depends on the contextual requirements for the speaker to signal control over ontology – i.e. speakers are aware of the way the content of different utterances is related and structured in the discourse. Not adhering to Maximize Presupposition in textual environments in which the requirements to signal control over ontology are high, e.g. in narratives, is claimed to lead to utterances detached from the preceding discourse and the undertone that the speaker did not plan their narrative beforehand. It is not clear, however, to what extent Eckhardt and Fränkel (2012)’s theory applies to communication outside of text production. For example, the extent to which speakers have to signal control over ontology in conversational dialogues remains to be investigated.

**Social convergence/divergence**

Another possible function of additives that has not been investigated yet – at least to our knowledge – is social convergence/divergence. More specifically, speakers who utter *too* may not only signal similarity between their utterance and the previously mentioned antecedent utterance but may also draw parallels between themselves and the antecedent speaker. This line of reasoning builds on Winterstein and Zeevat (2012)’s account on argumentative identity, which we introduced earlier, explaining the occurrence of additives with referring to the argumentative direction of antecedent and host: using *too* may signal that both antecedent speaker and the host speaker argue towards the same conversational goal.

Social convergence/divergence was introduced within the framework of Communication Accommodation Theory (e.g. Giles, 1973, 1977, 1979). The account is built on the assumptions that interlocutors negotiate personal and social identities through linguistic, discursive and non-linguistic tools (Gallois, Ogay & Giles, 2005). Thus, interlocutors may adjust their way of communicating to appear more like (converge) or distinct (diverge) from other interlocutors. A possible motive for speakers to converge their communicative behaviour could be the desire for approval (Gallois et al., 2005). In contrast, divergence may result from the wish to reinforce one’s social identity and/or the wish to dissociate themselves from other interlocutors. Whereas converging speakers are most
commonly received as cooperative, the reception of diverging speakers may vary. Divergence may be evaluated negatively, especially when diverging is judged to threaten social norms, e.g. politeness norms. However, diverging may likewise be received positively when diverging from a speaker entails converging with other interlocutors who share a common, positively valued group membership (Giles & Hewstone, 1982).

Converging and diverging are very much intertwined with social norms such as politeness which is construed within Politeness Theory as the interlocutors’ aim to maintain their positive or 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, speakers may aim to converge – e.g. by choosing similar topics as the other interlocutors – to maintain their positive face. In contrast, diverging from someone – e.g. by changing the topic – may maintain the speaker’s negative face but may likewise threaten their positive face, since diverging can come across as impolite.

Applied to additive particles, uttering *too* seems not only cooperative on the ontological level but also means that speakers converge with the antecedent speaker by signalling similarity between their utterances. This convergence move may come across as particularly strong if both antecedent utterance and host utterance convey values or opinions of both speakers, as in (69). In line with Communication Accommodation Theory, the motive of uttering *too* instead of leaving it out may reflect a general tendency to converge with speakers that share the same values (Gallois et al., 2005). However, one could also say that convergence is just a by-product of uttering *too* in obligatory environments. In this sense, uttering *too* is simply the easiest option to avoid making a controversial or face-threatening discourse move.

(69)  
a. Antecedent Speaker: I voted for the Green party.  
b. Host Speaker: I voted for them, too.

The derivation from the default of uttering *too* may be an instance of divergence. For example, one could assume that speakers omit *too* and avoid acknowledging the parallelism between their utterance and the antecedent utterance in order to distance themselves from the antecedent speaker. A possible motive for diverging from the antecedent speaker could involve contrasting values/opinion or that the antecedent’s behaviour threatens politeness norms. In this sense, speakers may wish to maintain their own negative face by not being associated with the antecedent speaker.\(^4\) In our

\(^4\)Not uttering *too* in (69b) may threaten the host speakers’ positive face, since diverging is seen as impolite overall and impolite behaviour may threaten the host speaker’s positive face. A motive for not uttering *too* may be that the
study, we aim to explore whether omitting additives in potentially obligatory environments can be understood as a means to diverge from an interlocutor’s turn.

5.4 The present study

This paper investigates under which circumstances speakers produce additive particles such as *too* when the antecedent turn in the discourse context varies along three dimensions: Similarity, Turn Distance and Politeness. Whereas Similarity has already been investigated and found to influence the production of additives, Turn Distance has not been empirically investigated yet. More specifically, whereas Kripke (2009) claims that a suitable antecedent merely has to be part of the interlocutors’ active context and does not necessarily have to be part of the immediately preceding utterance, the obligatoriness of additives is mostly empirically investigated by focusing on cases when the antecedent immediately precedes or is part of the host sentence. Additionally, we test the influence of Politeness on additive production to see whether intentionally omitting additives can be seen as a speaker’s attempt to diverge from an impolite antecedent speaker.

For this purpose, participants were asked virtually to attend a fictional work dinner. Their task was to interact with their colleagues and a waiter based on visual cues. We manipulated Similarity by asking participants to formulate an utterance that either perfectly matched or did not perfectly match the utterance of a previous speaker (antecedent speaker). Turn Distance was manipulated by the participant’s turn either immediately following the antecedent speaker’s turn (0 intervening turns) or following after three intervening turns. In order to test divergence we tested two groups, one in which participants conversed with polite/neutral speakers, and one in which the antecedent speaker would be impolite. This study consists of two experiments: in experiment I participants were invited to produce free text responses; in experiment II, they selected from a set of options (informed by the results of experiment I). Data and material for both experiments as well as the pre-registrations can be found here: https://osf.io/az2uf/.

5.4.1 Experiment I

The first experiment investigates the production of additives in an open choice paradigm in which participants were asked to type in their response in a text box. We coded participants’ responses as either containing or not containing any of the following additive presupposition triggers: *too, also* and *as well*. First, experiment I tests whether our experiment set-up elicits the use of additives and which additives participants would produce most frequently. We analyse the speaker’s binary antecedent speaker behaved impolite prior to uttering *(69a)* and the host speaker does not want to be associated with the antecedent speaker.
choice to include or omit an additive particle (dependent variable) to see how that choice is affected by Similarity, Turns, and Politeness. In experiment II we used a forced-choice paradigm based on the responses of experiment I.

**Participants**

Participants (N=78) 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 18 minutes). The age of the participants ranged from 18 to 83 years, with a mean of 37 years (median = 32, mode = 31). 40 participants stated their preferred pronoun as she/her, 37 chose he/him, and 1 chose they/them.

**Design and Materials**

Participants were asked to engage with their colleagues and a waiter by typing their responses in a text-field. Half of the critical items (‘order’ items) were about ordering dishes/drinks from a waiter, and half of the critical items (‘talk’ items) were about content related to work or daily life, e.g. the date and time of a meeting. Each critical item started with an introduction to the upcoming conversation; participants were presented with context pictures together with a question about their content. See Figure 5.1a for an ‘order’ item picture (always a menu) and Figure 5.1b for a ‘talk’ item picture (here a planner). In this way, participants were able to familiarise themselves with the visual cues (the menu/planner) they would later need for formulating their response. For example, the ‘order’ item context already prepared participants that they would be asked to order a glass of Chardonnay when it was their turn to order. The visual cues were then presented to participants again (Menu with ‘Chardonnay’ highlighted in Figure 5.3). This set-up is an attempt to mimic actual conversation, since interlocutors in an actual conversation do already have some knowledge about what they can contribute about a given topic when listening to the contributions of other interlocutors. It also allowed us to constrain the participant’s message content so that we could test what factors affected how they formulate an utterance about that content.
After having answered the context question, the conversation started. Each picture had a headline reinforcing the topic of the conversation. For ‘order’ items, the headlines would signify what interlocutors were ordering, e.g. ‘Ordering drinks’; for ‘talk’ items the headline would say what the conversation was about, e.g. ‘Discussion about the time of next week’s meeting’. Each conversation started with Omar’s turn followed by Lee, Amber and Robert, see Figure 5.2 for the first three turns. Each conversational turn was displayed separately and participants had to click a button to proceed to the next one. The last turn (Robert) was always displayed together with the visual cue and together with the request for the participants to write their response, see Figures 5.3a and 5.3b.

It is important to stress, that we tried to create scenarios where different speakers contribute to some conversational goal. We attempted to create a naturalistic scenario that are not necessarily list-like. However, one could argue that ordering food/drinks at a restaurant is pretty close to a list, in fact a list of orders. We nonetheless claim that by also testing more conversational-like scenarios
with the ‘talk’ items we were able to look at a wider range of constructions not just lists. In the ‘talk’ items the participants would contribute to a conversation about when a meeting is scheduled, or what cuisine is their favourite cuisine.

Turn Distance was manipulated such that the antecedent utterance either immediately preceded the participants’ turn (0 turn distance, see Figure 5.2a), or the antecedent and participants’ turn were intervened by the remaining colleagues (3 turn distance, see Figure 5.2b). Thus, the antecedent speaker for the 0 turn condition was always Robert and the antecedent speaker for the 3 turn condition was always Omar.

To manipulate the factor Similarity, the picture cues either encouraged participants to give a response that was perfectly similar or only vaguely similar to the antecedent utterance produced by either Robert or Omar. Taking the ‘order’ items as an example, the drink/dishes we asked participants to order either completely corresponded (perfect similarity) or did not correspond/only partially corresponded to the antecedent order (reduced similarity). For example, in Figure 5.3, participants were encouraged to order Chardonnay which perfectly matched Robert’s or Omar’s order. For the reduced similarity condition the participants’ order would have been Pinot Grigio which did not match Robert’s or Omar’s order. This was slightly different for orders in which participants were asked to order two dishes: in the perfect similarity condition, both the highlighted starter and main matched the antecedent order, as opposed to the reduced similarity condition in which only the starter but not the main matched the antecedent order (partial match).5 We will briefly discuss how these different kinds of dissimilarities influenced the participants’ use of additives in Section 5.4.1. We have illustrated the similarity manipulation for the ‘talk’ items in Appendix D.

---

5This manipulation was the same for the ‘talk’ items. Either both utterances were completely dissimilar (different favourite cuisines) or matched partially (same date but different month). In total, half of the critical items were manipulated to be dissimilar to the antecedent utterance and the other half to be a partially match.
(a) One intervening turn.

(b) Three intervening turns.

Figure 5.2: First three turns for an ‘order’ item (experiments I and II [6–7]) in which the participant will order Chardonnay across the two Distance conditions: (a) zero versus (b) three intervening turns between the antecedent speaker and the participant.
(a) One intervening turn.  
(b) Three intervening turns.

Figure 5.3: Last turn for the ‘order’ items (experiments I and II [6–7]) in the perfect similarity and neutral politeness condition (‘Chardonnay’ is highlighted). In the reduced similarity condition ‘Pinot Grigio’ would be highlighted.

We additionally investigated social divergence and manipulated the way the antecedent speakers phrased their orders: either neutrally, see Figure 5.3a, or impolitely, see Figure 5.4. Participants were randomly assigned to either the neutral or impolite condition. In the neutral condition, all four colleagues behaved politely, in contrast, in the impolite condition the two antecedent speakers (Omar and Robert) would make rude remarks towards the waiter or colleagues. However, their rude behaviour was restricted to the critical items and the attention checks stayed the same for both conditions. This resulted in a $2 \times 2 \times 2$ design, with two within-subjects factors (Similarity and Turn Distance) and one between-subjects factor (Politeness).
The 8 attention checks were similar to the ‘talk’ items. Participants were introduced to the visual cue, see Figure 5.5, followed by the colleagues’ turns and their own turn, see Figure 5.6. The only difference was that the colleagues’ turns were not manipulated to induce participants to use additive particles. Participants who failed to answer the context question incorrectly, and also failed to base their conversational contribution on the visual cue, were considered as having failed the attention check.

The experiment consisted of four blocks, each containing two attention checks, one ‘talk’ item, and one ‘order’ item (block1: drinks, block2: appetizer & main, block3: dessert, block4: drinks). Thus, each participant saw 8 critical items. Although the item order within each block was randomised, the order of the blocks was always the same.6

6In order to see whether block order influenced the probability to produce additives (i.e. whether the participants became more or less likely to produce additives as the experiment went on) we conducted models with block order as a predictor for experiments I and II. We did not find such a trend. The only weak trend that the order effects model revealed is that the frequency of choosing additives drops for Block3 in comparison to the other blocks which suggests that the items of Block3 deviated from the rest of the items: see Appendix D for more information.
What is the name of the gym that you are a member of?

Please respond here:

Figure 5.5: Attention check context for experiments I and II [6–7].
Figure 5.6: Attention check for experiments I and II [6–7].

Procedure

First, participants were asked to give informed consent to participate in the task. We then informed them about the structure of the experiment which was as follows: (1) engaging with different items of information (e.g. a to-do list) and interacting with fictional colleagues and a waiter, and (2)
a demographic questionnaire. After that we introduced participants to the task and their four colleagues. We explained that each conversation would start with them being asked to answer a question about visual material (e.g. a menu) that is related to the upcoming conversation. Participants were instructed to follow the conversation and type their own conversational contribution when asked to based on the visual material (e.g. a menu with highlighted dishes) that then will pop up again. We emphasised that there was no need to memorise anything at the beginning. Furthermore, we told participants that we are interested in the way they engage with others and asked participants to write their responses in such a way that they reflect their preferred way of speaking. The experiment was followed by the voluntary demographic questionnaire.

Predictions

Our predictions are based on the previously discussed theoretical accounts and experimental studies. Overall, following the Maximize Presupposition principle, participants are predicted to most frequently use additives for highly similar and recent antecedents, as this comes closest to what has been empirically investigated as being an obligatory context.

Similarity  With increasing similarity between antecedent and host sentence, the preference to use additives should increase. Thus, participants are predicted to produce additives more in the perfect similarity condition than in the reduced similarity condition.

Turn Distance  With increasing turn distance between antecedent and host sentence, the preference to use additives should decrease. Hence, participants are predicted to produce additives less in the 3 turns condition (three intervening turns) than in the 0 turns condition (zero intervening turns).

Politeness  When antecedent speakers behave impolitely, speakers should be more likely to drop the use of additives to diverge from the antecedent speakers. Thus, participants are predicted to produce additives less in the impolite condition than in the neutral condition.

Interactions  We will also explore the interactions of these three factors. This part will be exploratory and inform our second experiment.

Analysis

We coded participants’ responses for both experiments as either containing (1) or not containing (0) any of the following additive presupposition triggers: too, also and as well. The data was analysed fitting a Bayesian logistic 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 factors Similarity, Turn Distance, Politeness and their interactions were included to predict the probability to produce additives. All three factors were sum-coded, whereby the levels perfect, zero intervening turns and neutral politeness were the reference categories coded with 1 (the other levels were coded with -1). The model included varying intercepts and slopes for participants and items, assuming that the effects of Similarity, Turn Distance and their interaction vary between participants, and the effects of Similarity, Turn Distance, Politeness and their interactions vary between items.

We used weakly regularising priors, which allowed a reasonably wide range of parameter values and penalised very extreme values. The priors for the by-expression intercepts were normal distributions with mean 0 and standard deviation 3. 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 et al., 2016).

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 whereby the first 1000 iterations were disregarded as part of the warm-up phase leading to 12000 iterations available for analysis.

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 et al., 2012).

---

7We had pre-registered both experiments with treatment coding but used sum-coding instead. This was due to the fact that (i) sum-coding helped the models to converge (especially for experiment I), (ii) using sum-coding the associations of parameters correspond intuitively to a main effect of a IV being the average effect of changing between levels of the IV (Levy, 2014).

8In alignment with our preregistration, we ran two models with more uninformative/wider priors, see Appendix D for details. Applying wider priors yielded coefficients with wide standard deviations and credible intervals. One of the model outputs suggests that the chains have not mixed well. This finding suggests that our data was sparse and reiterates that it is crucial to rerun the study with more participants which we did by running experiment II.
Results

Participants were above 96% accurate for the attention checks, which suggests that they paid attention during the experiment. Overall, participants used additive formulations for target items throughout the experiment (in total 71 times out of 624), and thus, 11% of the participants’ utterances contained additives, among them too, also, and as well, with too being the most frequent choice, see Table 5.1 for more details.

<table>
<thead>
<tr>
<th>Additive</th>
<th>Item Type</th>
<th>Absolute Freq.</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>too</td>
<td>Order Item</td>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>too</td>
<td>Talk Item</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>also</td>
<td>Order Item</td>
<td>15</td>
<td>26</td>
</tr>
<tr>
<td>also</td>
<td>Talk Item</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>as well</td>
<td>Order Item</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>as well</td>
<td>Talk Item</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Raw additive counts of experiment I [6] for each item type and overall.

There was a small difference in how frequently additives were used within reduced similarity items. Recall that half of the critical items were manipulated to be a categorical mismatch between antecedent and participant utterance, and the other half was manipulated to be a partial match. Participants would still occasionally use additives when there was a partial match as opposed to no match at all.
As Figure 5.7 illustrates, additives were most frequently used when host utterance and antecedent utterance were similar and when there were zero intervening turns. Frequency was reduced when three turns intervened. In contrast, the difference between the neutral and impolite conditions is less clear.
Table 5.2: Experiment I [6]: Population-level estimates of the logistic regression model in log-odds with the standard errors and 95% credible intervals. In the table the intercept (i.e. the grand mean) is listed first, then the estimates for Similarity, Turn Distance, Politeness and their interactions follow. Similarity is the change in log-odds for perfect similarity (-1 reduced, 1 perfect), Turn Distance is the change in log-odds for no turn (-1 three turns, 1 no turns), and Politeness is the change in log-odds for a neutral antecedent speaker (-1 neutral, 1 impolite). \( \hat{R} \) is a convergence diagnostic which compares the between- and within-chain estimates. Values larger than 1 suggest that the chains have not mixed well.

These observations are supported by the outcome of the analysis. Averaging over Turn Distance and Politeness, a change to the perfect similarity condition meant an increase in log-odds of additive production \( (\hat{\beta} = 1.58, \text{CrI:} [0.78, 2.42]) \). Similarly, averaging over Similarity and Politeness, a change to the zero turns condition meant an increase in log-odds of additive production \( (\hat{\beta} = 1.17, \text{CrI:} [0.51, 1.93]) \). Both effects seem reliable, since the credible intervals of both effects do not include zero. However, we did not find reliable evidence that the antecedent speakers’ politeness influenced additive production. In Figure 5.8, we illustrated the predicted probabilities for speakers to produce additives given our data and model. Speakers are predicted to produce additives most frequently for perfect similarity and when no turns intervene between their utterance and the antecedent utterance and least frequently for reduced similarity when three turns intervene.
Figure 5.8: Experiment I [6]: Predicted probabilities to produce additive particles given the model and the data for perfect (left) and reduced similarity (right) and 0 and 3 intervening turns. The figure contrasts the probability to produce additives when faced with a neutral antecedent speaker (blue) with an impolite antecedent speaker (red).

The credible intervals of all interaction terms include zero which suggests that our three predictors did not interact in a systematic way. We compared the expected log predictive density of the full interaction model to models with reduced interaction terms, a main effects only model and a model that only included Similarity and Turn Distance as predictors. Model comparison was carried out via PSIS-LOO approximation (Pareto smoothed importance sampling leave-one-out approximation; Vehtari and Gelman (2015); Vehtari et al. (2017)). The approximation disclosed that the model including the predictors Similarity and Turn Distance and their interaction without the predictor Politeness had the highest predictive accuracy, followed by a model that included all three predictors as main effects. The model with three way interaction which we introduced above had the lowest predictive strength. However, the differences between the models were small, see Table 5.3.
Table 5.3: Experiment I [6]: The table shows the difference in the models’ expected predictive accuracy (elpd_diff) and the standard error of the difference in elpd (se_diff) whereby the preferred model is listed first.

<table>
<thead>
<tr>
<th>Similarity * Turn Distance</th>
<th>elpd_diff</th>
<th>se_diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sim + TD + Politeness</td>
<td>-1.5</td>
<td>2.9</td>
</tr>
<tr>
<td>Sim * P + TD</td>
<td>-2.1</td>
<td>3.5</td>
</tr>
<tr>
<td>Sim + TD * P</td>
<td>-3.5</td>
<td>3.1</td>
</tr>
<tr>
<td>Sim * TD + P</td>
<td>-3.9</td>
<td>2.4</td>
</tr>
<tr>
<td>Sim * TD * P</td>
<td>-4.5</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Interim Discussion

Results of experiment I suggest that speakers use additives more frequently for perfect similarity than reduced similarity and when the antecedent turn directly precedes the speakers’ turn as opposed to when it does not. There was no reliable evidence that speakers additive production is influenced by the antecedent speaker’s politeness.

The overall use of additives was rather low (approx. 11%), which might mean that additives are less favoured than expected by the Maximize Presupposition principle. Recall that according to the Maximize Presupposition principle additives are obligatory whenever there is a suitable antecedent. Thus, at least for the condition with perfect similarity and zero intervening turns participants should have used additives much more frequently. One may argue that the reason for the low frequency in additive use is due to our experimental design not being close enough to actual production. However, in comparison to previous acceptability/felicity judgement tasks (e.g. Amsili & Beyssade, 2009; Grubic & Wierzba, 2019) or text production tasks (Eckhardt & Fränkel, 2012) our experiment comes closer to actual interaction.

A possible reason for the lower-than-expected occurrence of presupposition triggers was that participants used the formulation same (as) frequently to signal similarity between their and the antecedent utterance, see (70)–(71) for examples. Since same (as) is not a presupposition trigger, we did not include it in our analysis. Nevertheless, we will briefly discuss in what way speakers used same (as) in our experiment.

(70) ‘Order’ item examples
  a. The same for me please!
  b. Can I get the same as omar please
(71) ‘Talk’ item examples
   a. I did the same Robert.
   b. Omar, I’ve done the same as you! £15 and a bottle of wine.

In Figure 5.9 we plotted the absolute frequencies of each additive particle and the formulation *same (as)*. The absolute frequencies suggest that, in contrast to additive particles, *same (as)* was more frequently used when the antecedent speaker was impolite as opposed to when the antecedent speaker was neutral. A possible reason as to why speakers who aim to acknowledge similarities between their utterance and the utterance of an impolite antecedent speaker do so by choosing *same (as)* more frequently than any of the additive particles may be that speakers produce *same (as)* do not consider themselves to converge with the antecedent speaker to the same extent. More specifically, one could argue that *same (as)* highlights a different kind of similarity in that it only refers to what the antecedent speaker ordered (‘I'd like the same as Robert.’) or what the antecedent speaker did (‘I did the same as you, Omar.’). In contrast, producing additives might involve a greater sense of alliance with the antecedent speaker since it seems to affect the whole proposition that is expressed by the utterance including the personal pronoun (‘I got the wine and the tie, too.’/ ‘I also got the wine and the tie.’). However, these are just post-hoc speculations and need to be investigated further in future studies.
Figure 5.9: Raw counts for additives and *same* of experiment I [6], contrasting politeness (neutral is green, impolite is pink). ‘PerfectSim’ stands for perfect similarity, ‘ReducedSim’ for reduced similarity, ‘0Turns’ for the 0 turns distance condition, and ‘3Turns’ for the 3 turns distance condition.

### 5.4.2 Experiment II

The second experiment investigates the use of additives in a forced-choice paradigm in which participants can choose between the bare assertion (72), the assertion plus additive (73), an incorrect response (74) and the option to type their own response (‘other’).

1. I’d like a glass of Chardonnay, please.
2. I’d like a glass of Chardonnay too, please.
3. I’d like a Martini, please.

We decided to focus on *too*, since it was the most frequently chosen option in experiment I, and we chose to investigate only one additive under the assumption that *too, also* and *as well* are used similarly (at least in our experimental setting). However, we included the participants’ formulated utterances (option ‘other’) in our analysis. Similar to the pre-determined utterance choices we coded participants’ ‘other’ responses as either containing an additive (1) or not containing an additive (0). Thus, although we restrict the participants’ pre-determined choices to use *too*, the option ‘other’
opens up the possibility to use any other additive.

The experimental design of experiment II more closely resembles acceptability judgement tasks which are frequently conducted to analyse the obligatoriness of additives: participants see the direct contrast between antecedent utterance and an utterance with or without an additive. Seeing both utterance options we expect more frequent use of additives. The aim of experiment II is to replicate the effects of Similarity and Turn Distance and to shed more light on interaction effects by increasing the number of participants as well as the expected number of additive-containing continuations. In addition to investigating Similarity as a binary factor, we will take an exploratory look at the distinction between partial similarity and no similarity, based on the results of experiment I.

Participants

Participants (N=140) 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 12 minutes). The age of the participants ranged from 18 to 75 years, with a mean of 36 years (median = 34, mode = 32). 3 participants stated their preferred pronoun as they/them, 16 participants as he/him and 122 as she/her.

Design and Materials

We used the same experimental design as for experiment I, see Section 5.4.1. The only difference was that participants were asked to engage with their colleagues and a waiter by either choosing one of the three utterance choices or by typing their own response (choice ‘other’), see (75–77). Crucially, we presented different utterance choices for partial similarity items (76) as opposed to no similarity items (77). This decision was based on the results of experiment I which showed that participants still produced additives for partial similarity items as opposed to no similarity items. In these instances participants used additives in constructions such as (76b). We decided to use this formulation instead of ‘I gave £15 and some chocolate, too’, since we did not want to gloss over the fact that too can still felicitously be used for parts of the antecedent utterance in the reduced similarity condition. Being aware of an enforced difference between partial and no similarity we explored the factor Similarity as a three level factor in a second analysis.

(75) Perfect Similarity (antecedent speaker: glass of Chardonnay)
   a. I’d like a glass of Chardonnay, please.
   b. I’d like a glass of Chardonnay too, please.
   c. I’d like a Cosmopolitan, please.
(76) Partial Similarity (antecedent speaker: £15 and wine)
   a. I gave £15 and some chocolate.
   b. I gave £15 too but bought him some chocolate instead.
   c. I gave £25 and whisky.

(77) No Similarity (antecedent speaker: a pint of Heineken)
   a. I’d like a pint of Stella Artois, please.
   b. I’d like a pint of Stella Artois too, please.
   c. I’d like a Martini, please.

Procedure

The same procedure as in experiment I was followed: (1) informed consent, (2) experiment, (3) questionnaire, see Section 5.4.1 for more details. The only difference was that we added a question to the questionnaire about whether they thought that the utterance choices reflected their way of speaking. Participants could respond by adjusting a slider from 0 ('Definitely not') to 100 ('Definitely yes'). The experiment consisted of four blocks that were presented in the same order to all participants.

Predictions

Again, following the Maximize Presupposition principle, participants are predicted to most frequently use additives for highly similar and recent antecedents, since it comes closest to what has been empirically investigated as being an obligatory context. The remaining predictions are based on the results of experiment I.

Similarity We predict that participants choose additives more in the perfect similarity condition than in the reduced similarity condition.

Turn Distance We predict that participants choose additives less in the 3 turns condition (three intervening turns) than in the 0 turns condition (zero intervening turns).

Politeness We predict that participants choose additives less in the impolite condition than in the neutral condition.

Interactions We will further explore the interactions of these three factors. This part will again be exploratory, since the data of experiment I was too sparse to inform any concrete predictions here.
Analysis

We analysed the data as in experiment I, by fitting a Bayesian logistic regression model with varying by-item and by-participants intercepts and slopes. The experimental factors Similarity, Turn Distance and Politeness and their interactions were included to predict the probability to use *too*. All factors were sum-coded as in the analysis of experiment I. The model included varying intercepts and slopes for participants and items, assuming that the effects of Similarity, Turn Distance and their interaction vary between participants, and the effects of Similarity, Turn Distance, Politeness and their interactions vary between items. The choice of priors and the sampling process were the same as for the experiment I analysis, see Section 5.4.1 for a detailed description.9

Results

All participants were above 97% accurate for the attention checks, which suggests that they paid attention during the experiment. Overall, participants used additive formulations throughout the experiment and more often than in experiment I: 32.62% of the participants’ utterances contained additives. ‘Order’ items elicit again more uses of additives (200) than ‘talk’ items (168). On average participants rated the naturalness of the pre-determined utterance choices with 79 (the measure ranged from 0 to 100). Correspondingly, the option ‘other’ was chosen 73 times. After inspection of the ‘other’ response, 12 instances were included that contained additives of which *too* was used 9 times, *as well* 2 times and *also* 1 time. ‘Other’ responses containing *too* often entailed further interaction with either a neutral (78a) or impolite (78b–78c) antecedent speaker.

(78)  a. Great minds think alike Robert. I’ll have the sticky toffe pudding and an expresso too.
     b. i’d like a pin of heineken too please. Robert you shouldn’t speak to the waiter in that way it is disrespectful
     c. To waiter: The Breaded brie and king prawn linguine for me too, please. To Omar (once the waiter has left): Omar, are you okay? I think you might have been a bit abrupt with the waiter and that’s not like you.

Figure 5.10 illustrates that, similarly to experiment I, additives were most frequently used when host utterance and antecedent utterance were similar and when there were zero intervening turns. In contrast to experiment I, participants used additives in the reduced similarity/three intervening turns condition. Moreover, the frequency of additive use seems to be less when speaking after an impolite speaker than to a neutral speaker for both perfect similarity conditions and in the reduced

---

9In alignment with our preregistration, we ran two models with more uninformative/wider priors yielding similar posteriors, see Appendix D for details. Overall, the models with wider prior distributions yielded higher posterior means but also introduced more uncertainty, specifically for the posterior distributions of the interaction coefficients.
similarity with three intervening turns. The opposite seems to be the case for the reduced similarity condition with zero intervening turns.

![Figure 5.10: Experiment II [7]: Relative frequency of producing additives. Contrasting 0 (left) versus 3 (right) intervening turns, perfect versus reduced similarity (x-axis) and speaking after a neutral (blue) versus an impolite (red) antecedent speaker.](image)

These observations are supported by the outcome of the analysis, see Table 5.4. Holding everything else constant, perfect similarity between antecedent utterance and host utterance led to an increase in log-odds of additive production ($\hat{\beta} = 1.30$, CrI:[0.53, 2.00]), as well as when the antecedent utterance directly preceded the host utterance ($\hat{\beta} = 0.74$, CrI:[0.34, 1.15]), and when speaking after a neutral antecedent speaker ($\hat{\beta} = 0.36$, CrI:[0.05, 0.69]). Overall, Similarity seems to affect the production of additives to a greater extent than Turn Distance and Politeness. The credible intervals of all effects do not include zero.
Table 5.4: Experiment II [7]: Population-level estimates of the logistic regression model in log-odds with the standard errors and 95% credible intervals. In the table the intercept (i.e. grand mean) is listed first, then the estimates for Similarity, Turn Distance, Politeness and their interactions follow. Similarity is the change in log-odds for perfect similarity (-1 reduced, 1 perfect), Turn Distance is the change in log-odds for no turn (-1 three turns, 1 no turns), and Politeness is the change in log-odds for a neutral antecedent speaker (-1 neutral, 1 impolite). $\hat{R}$ is a convergence diagnostic which compares the between- and within-chain estimates. Values larger than 1 suggest that the chains have not mixed well.

The credible intervals of all interaction terms include zero. The most reliable interaction coefficient seems to be the three way interaction coefficient ($\hat{\beta} = 0.24$, CrI:[-0.02, 0.50]) predicting an increase in log-odds for perfect similarity, zero intervening turns and a neutral antecedent speaker. We compared the expected log predictive density of the full interaction model to models with reduced interaction terms, a main effects only model and a model that only included Similarity and Turn Distance and their interaction as predictors. Model comparison was carried out via PSIS-LOO approximation [Vehtari and Gelman (2015); Vehtari et al. (2017)] which disclosed that the model including all predictors and their interactions had the highest predictive accuracy, followed by a model that included Similarity and Turn Distance and their interaction plus Politeness as main effect. The model only including Similarity and Turn Distance had the lowest predictive accuracy. However, the difference between the models is very small, see Table 5.5.

<table>
<thead>
<tr>
<th>Coefficient Estimates</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>$\hat{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.39</td>
<td>0.33</td>
<td>-2.07</td>
<td>-0.75</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity</td>
<td>1.30</td>
<td>0.37</td>
<td>0.53</td>
<td>2.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Turn Distance</td>
<td>0.74</td>
<td>0.21</td>
<td>0.34</td>
<td>1.15</td>
<td>1.00</td>
</tr>
<tr>
<td>Politeness</td>
<td>0.36</td>
<td>0.16</td>
<td>0.05</td>
<td>0.69</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity*Turn Distance</td>
<td>-0.16</td>
<td>0.18</td>
<td>-0.51</td>
<td>0.21</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity*Politeness</td>
<td>-0.00</td>
<td>0.15</td>
<td>-0.29</td>
<td>0.29</td>
<td>1.00</td>
</tr>
<tr>
<td>Turn Distance*Politeness</td>
<td>-0.24</td>
<td>0.18</td>
<td>-0.60</td>
<td>0.08</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity<em>Turn Distance</em>Politeness</td>
<td>0.24</td>
<td>0.13</td>
<td>-0.02</td>
<td>0.50</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 5.5: Experiment II [7]: The table shows the difference in the models’ expected predictive accuracy (elpd_diff) and the standard error of the difference in elpd (se_diff) whereby the preferred model is listed first.

In Figure 5.11, we illustrated the predicted probabilities for speakers to utter additives given our data and model. Speakers are predicted to produce additives more frequently for perfect similarity and when no turns intervene between their utterance and the antecedent utterance and least frequently for reduced similarity when three turns intervene. There seems to be a more pronounced difference between neutral (blue) and impolite conditions (red) than in experiment I for the perfect similarity conditions such that speakers are predicted to produce additives more frequently when talking to a neutral antecedent speaker as opposed to an impolite antecedent speaker.

![Predicted probabilities of Additive production](image)

Figure 5.11: Predicted probabilities to produce additive particles given the model and the data of experiment II [7] for perfect (left) and reduced similarity (right) and 0 and 3 intervening turns. The figure contrasts the probability to produce additives when faced with a neutral antecedent speaker (blue) with an impolite antecedent speaker (red).
As was mentioned above, we ran a further analysis in which we treat the factor Similarity as a three level predictor (perfect similarity, partial similarity, no similarity). Recall that the level reduced similarity combined partial and no similarity items. We re-ran the same model as before only this time with Similarity as a three level factor which was sum-coded, see Table 5.6.

<table>
<thead>
<tr>
<th>Similarity1</th>
<th>Similarity2</th>
</tr>
</thead>
<tbody>
<tr>
<td>perfect</td>
<td>1</td>
</tr>
<tr>
<td>dissimilar</td>
<td>0</td>
</tr>
<tr>
<td>partial</td>
<td>-1</td>
</tr>
</tbody>
</table>

Table 5.6: Experiment II [7]: Contrast coding for the three-level Similarity factor.

<table>
<thead>
<tr>
<th>Coefficient Estimates</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>( \hat{R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.06</td>
<td>0.34</td>
<td>-2.71</td>
<td>-1.38</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity1</td>
<td>1.91</td>
<td>0.40</td>
<td>1.08</td>
<td>2.67</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity2</td>
<td>-2.06</td>
<td>0.48</td>
<td>-2.95</td>
<td>-1.07</td>
<td>1.00</td>
</tr>
<tr>
<td>Turn Distance</td>
<td>0.78</td>
<td>0.26</td>
<td>0.28</td>
<td>1.31</td>
<td>1.00</td>
</tr>
<tr>
<td>Politeness</td>
<td>0.38</td>
<td>0.21</td>
<td>-0.03</td>
<td>0.82</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity1*Turn Distance</td>
<td>-0.21</td>
<td>0.25</td>
<td>-0.71</td>
<td>0.30</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity2*Turn Distance</td>
<td>-0.25</td>
<td>0.36</td>
<td>-0.93</td>
<td>0.49</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity1*Politeness</td>
<td>-0.03</td>
<td>0.23</td>
<td>-0.48</td>
<td>0.43</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity2*Politeness</td>
<td>0.13</td>
<td>0.34</td>
<td>-0.52</td>
<td>0.81</td>
<td>1.00</td>
</tr>
<tr>
<td>TurnDistance1*Politeness</td>
<td>-0.35</td>
<td>0.24</td>
<td>-0.85</td>
<td>0.12</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity1<em>Turn Distance</em>Politeness</td>
<td>0.34</td>
<td>0.23</td>
<td>-0.12</td>
<td>0.80</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity2<em>Turn Distance</em>Politeness</td>
<td>-0.27</td>
<td>0.36</td>
<td>-1.01</td>
<td>0.42</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 5.7: Experiment II [7]: Logistic regression output with Similarity as a three-level factor. Population-level estimates of the logistic regression model in log-odds with the standard errors and 95% credible intervals. In the table the intercept is listed first, then the estimates for Similarity, Turn Distance, Politeness and their interactions follow. \( \hat{R} \) is a convergence diagnostic which compares the between- and within-chain estimates. Values larger than 1 suggest that the chains have not mixed well.

The output suggests that, averaging over the other effects, perfect similarity led to an increase in log-odds of additive production (\( \hat{\beta} = 1.91 \), CrI:[1.08, 2.67]), and in contrast, dissimilarity led to a decrease in log-odds of additive production (\( \hat{\beta} = -2.06 \), CrI:[-2.95, -1.07]). Partial similarity also led to a decrease in additive production but less so than dissimilarity (approximate decrease: -1.91 in log odds).
Interim Discussion

We were able to replicate parts of the findings of experiment I: speakers seem to use additives more for perfect similarity than reduced similarity and when the antecedent turn directly precedes the speakers’ turn as opposed to when it does not. Overall, the additive production was again lower than assumed in potentially obligatory environments (perfect similarity/ zero intervening turns condition). Furthermore, in contrast to experiment II we found evidence that speakers more frequently use additives when speaking after a neutral speaker and more frequently refrain from producing additives when the antecedent speaker is impolite.

(79) Partial Similarity (antecedent speaker: £15 and wine)
   a. I gave £15 and some chocolate.
   b. I gave £15 too but bought him some chocolate instead.
   c. I gave £25 and whisky.

(80) No Similarity (antecedent speaker: a pint of Heineken)
   a. I’d like a pint of Stella Artois, please.
   b. I’d like a pint of Stella Artois too, please.
   c. I’d like a Martini, please.

By investigating Similarity as a three level factor we discovered differences between the way partial similarity and dissimilarity influenced participants’ additive use. Recall that we gave participants the option to choose between three utterances: an utterance with too, an utterance without too, and an incorrect response, see (79–80). Partial similarity and no similarity differ from each other in that too can still be felicitously used in the partial condition but not in the no similarity condition. We conjecture that the words but and instead in (79b) for partial similarity may already express a dissociation between what the antecedent speaker ordered/said so that dropping too is no longer necessary. However this part is speculative and has to be investigated further.

5.5 Discussion

5.5.1 Salience of the antecedent

Overall, our results on Similarity and Turn Distance helped to clarify (i) what material may serve as a suitable antecedent for additives to be produced felicitously, and (ii) whether additives are obligatory given the context entails a suitable antecedent. More specifically, our results on Turn Distance suggest that speakers use additives more frequently when their conversational turn immediately preceded the antecedent turn. These results shed more light on how immediately a suitable
antecedent has to precede a speakers’ utterance for additives to be used felicitously. Whereas it is true that immediately preceding antecedents did elicit a higher rate of additives, speakers still used additives for an antecedent separated by three intervening turns. Thus, in alignment with Kripke (2009), proximity does not necessarily determine whether additives can be used felicitously; however, proximity does seem to influence the salience of the antecedent which in turn influences whether speakers produce additives or not.

While we did not investigate the distinction between antecedents consisting of an individual plus property (Beaver & Zeevat, 2007) versus an individual (Geurts & van der Sandt, 2004) (see Section 5.3.2) – all our utterances contained an individual and their property (e.g. ‘I like Korean cuisine, too.’) – we found that speakers chose additives even for partial property matches and when three turns intervened between antecedent and host utterance. The results likewise align with Zeevat (2002)’s observation that presuppositions triggered by additives can be partially resolved. For example (81a), taken from Zeevat (2002, p. 63), only mentions eating and does not imply that a dinner is involved, and yet suffices as an antecedent for (81b).

(81)  
     a. Bill believes that Mary will eat in New York.  
     b. John is having dinner in New York, too.

Overall, additives were used less in the perfect similarity and zero intervening turns condition – which comes closest to what has been empirically investigated as being an obligatory context – than predicted by accounts postulating the obligatory use of additives. Even when the antecedent utterance directly preceded the speakers’ utterance and both were perfectly similar in formulation and content, the probability of speakers choosing the utterance containing too was on average only 0.6 (in experiment II). This suggests that while a salient antecedent favours the use of additives it does not lead to obligatory use of additives. In alignment with Amsili and Beyssade (2009)’s discussion of discourse structures that make the use of additives less obligatory, our findings suggest that additives may not be necessary if there are discursive cues – not necessarily as obvious as discourse structures such as lists – that make the similarity between the content of different utterance sufficiently obvious.10 Thus, our results align with those accounts which stressed that additives fulfil a discursive function such as discourse management.

10Even if one were to assume that ‘order’ items are list-like and assuming that the ‘talk’ items are not list-like, participants used additives more frequently in the more list-like scenario for both experiments. Thus, it seems reasonable to assume that there might be other discursive factors at play apart from the list environment.
5.5.2 Discourse organisation and grounding

We argue that the discourse-managing properties of additives – signalling similarities (e.g. Amsili et al., 2016; Winterstein & Zeevat, 2012) between antecedent and host utterance as well as ontological control (i.e. the awareness of how content is organised in the discourse) (Eckhardt & Fränkel, 2012) – seem interconnected with their presuppositional nature, part of which is anaphoric. We propose that both properties may make additives an important grounding tool for interlocutors. Grounding is understood as a process by which interlocutors collect and coordinate knowledge (Clark & Brennan, 1991). This process involves accumulating new knowledge and updating the interlocutors’ shared common ground with newly acquired information, while keeping track of what information has already entered the common ground. By using additives, speakers (i) refer back to content in the common ground, (ii) acknowledge parallelism between content of the common ground and their utterance, and (iii) signal that they kept track of what information is part of the common ground.

Furthermore, producing additives as part of the grounding process may come with a working memory cost, such that keeping track of what information is already shared knowledge and in the common ground may be more difficult with increasing conversational turns/time. Thus, not producing additives may happen unintentionally and may happen more if the turn with the suitable antecedent happened longer ago. The observed effect of Turn Distance seems to corroborate this: speakers more frequently used additives when their turn immediately preceded the antecedent turn. Furthermore, seeing additives as grounding tools may explain why additives were not always used when they were claimed to be obligatory: speakers might not always deem it necessary to explicitly signal their involvement in grounding via additive use.

5.5.3 Converging/diverging

Regarding the factor Politeness we found that speakers use additives more frequently when speaking after a neutral antecedent speaker and omit additives more frequently when speaking after an impolite antecedent speaker. Hence, the use of additives as a cooperative attempt to participate in the grounding process can be seen as a convergence move that happens by default. Disengaging from this process may be due to a speaker’s attempt to diverge from the antecedent speaker. Thus both Turn Distance and Politeness may account for instances when speakers choose not to produce additives. Whereas Turn Distance may explain unintentional omission of additives, Politeness may explain intentional omission of additives even with a highly salient (high similarity and close proximity) antecedent present.

A few responses in experiments I and II suggest that, for some participants, grounding had priority
over diverging from the antecedent speaker by omitting additives: after having formulated an utterance containing an additive, some participants continued to express their disagreement with the antecedent speaker’s behaviour in a second sentence, see (78). In this way participants remained cooperative towards the waiter and the other interlocutors. This way of dealing with an impolite speaker may be one of the reasons for why we did not find an effect of Politeness in experiment I. Since participants formulated their own utterances in experiment I, they were possibly more likely to follow this strategy than participants in experiment II who would most likely choose pre-formulated utterances.

Another explanation for our results would be to argue that the contribution of the antecedent speaker in the impolite scenario was less similar to the participants contribution (in the second experiment where the politeness effect was found) than in the polite scenario which is why participants might have omitted additives more frequently in the impolite scenario than in the polite scenario. For example, in the polite scenario the antecedent speaker’s contribution was ‘I’d like a glass of Chardonnay, please’ and in the impolite scenario ‘Although you served me warm white wine least time, I’d like a glass of Chardonnay.’ whereby the participants were asked to choose between ‘I’d like a glass of Chardonnay, please.’ or ‘I’d like a glass of Chardonnay too, please.’ which perfectly matches the polite but not the impolite antecedent speaker. We cannot exclude this possibility at this point.11

Overall, the effect of Politeness on the production of additives needs further investigation. While the different outcomes of experiments I and II may be due to sparse data in experiment I, replication studies are necessary not only to explore Politeness further but also to shed more light on the possible effect that open versus forced-choice tasks have on production data. Furthermore, we only investigated one potential reason for speakers to diverge from an antecedent speaker, namely when they behave impolitely. There may be other reasons with a potentially greater influence on whether speakers diverge or not such as when the antecedent speaker expounds controversial, false or immoral opinions.12 However, such reasons are difficult to test in an experimental setting due to ethical considerations.

---

11 We do not assume that the additive could have been perceived as ambiguous in the impolite scenario in which the antecedent speaker asserted two propositions (e.g. Although you served me warm white wine least time and I’d like a glass of Chardonnay) which theoretically could serve as antecedents. At least for the forced-choice paradigm, the utterance containing the additive disambiguated which one of the propositions asserted by the antecedent speaker could serve as a potential antecedent.

12 Imagining our restaurant scenario in which the antecedent speaker is clearly racist (and assuming the participant is not): would this disagreement be mirrored in a lack of additive usage by the participant when ordering the same food as their racist colleague?
5.5.4 Focus

A crucial property that we have not discussed in this paper is the association of additives with focus/contrastive topics. While we described additives as presupposing a propositional alternative, that alternative has been claimed to be determined by what constituent of the host sentence is focused (e.g. Rooth, 1985). For example, in (82a) *too* associates with the prosodically stressed *I* which makes the presupposed propositional alternative one that demands an alternative subject. In contrast, in (82b), *too* associates with ‘she’ which makes the propositional alternative one about who is starting in January.

(82) a. [I]$_F$ noted down that she’s due to start in January, too.
   presupposes: *Someone other than the speaker noted down that she’s due to start in January.*

b. I noted down that [she]$_F$’s due to start in January, too.
   presupposes: *The speaker noted down that someone other than ‘she’ is due to start in January.*

Since we did not provide participants with audio recordings in experiment II, they had to assign their focus individually, e.g. participant1 may have placed their focus on ‘I’ (82a) and participant2 on ‘she’ (82b). Triggering different presuppositions may have led to different assessments by participants of whether there was a suitable antecedent in the preceding discourse which would in turn induce some variation in additive use between participants. While this is possible, we argue that by presenting the overall discourse topic, we reduced such variation to some extent. Pragmatically, focused material generally addresses the explicit or implicit *Question under Discussion* (QUD). Following Roberts (1996), the QUD corresponds to the current discourse topic. Recall that in our study we set the discourse topic for each experimental item headlining each picture of a given conversation. For the item in example (82) the topic was ‘Discussion about when the new colleague is starting’. Such a topic would most likely be associated with a focus on ‘I’ (82a). Nonetheless, it remains to be investigated whether and how focus plays a role in the production of additives.

5.6 Conclusion

In this paper we investigated the circumstances under which additives are used. The findings of experiments I and II suggest that the production of additives is less obligatory than previously assumed: our findings show that increased similarity and close proximity between antecedent and host sentences favour increased use of additives but do not make their use obligatory. Furthermore, our findings suggest that omitting additives (even with a highly salient antecedent present) may be understood as a divergence move away from the antecedent speaker. However, this has to be
investigated further. Taking together additives’ presuppositional nature and discursive function, we introduced a model in which speakers use additives as a grounding tool to organise the discourse and maintain discourse coherence. Grounding itself could be understood as a cooperative process by which speakers converge with their interlocutors.
Summary
Chapter 6

Discussion

Part II investigated the production of know and its potential strategic advantage over (un)certainty expressions and the bare assertion. Taken together, the results of Part II seem to suggest that one of the motives for speakers to presuppose content by uttering know is to structure the discourse to their liking. In contrast, Part III suggests that one of the motives for speakers to presuppose content by uttering additives such as too is to participate actively in the interlocutors’ grounding process which involves managing the discourse. The following sections 6.1 and 6.2 will illustrate potential differences in the way factive and additive presupposition triggers interact with discourse factors.

Furthermore, while the methodology developed in this thesis illustrated ways in which the comprehension and production of pragmatic/semantic phenomena such as presuppositions can be investigated, there are potential ways in which the methodology could be expanded. Section 6.3 will review a few open questions and shortcomings of the methodology used in this thesis. Additionally, section 6.4 will talk about the relevance of investigating strategic communication for other areas of linguistics.

6.1 Differences between Presupposition Triggers

This thesis investigated possible motives that speakers may have to presuppose rather than assert content, focusing on different presupposition triggers. It was proposed that presupposition triggers are important tools for speakers to structure and manage the discourse. More specifically, Part II of this thesis suggested that speakers use factive presupposition triggers (i.e. know) as discourse structuring tools which involves projecting future discourse moves: speakers may presuppose content to make potentially controversial content less accessible. Furthermore, Part III argued that
additive particles such as *too* are discourse managing tools which involves indicating the state of the common ground: speakers utter additives to maintain discourse coherence and to prevent marking the discourse in an unusual way. I will argue that the distinct discursive functions of factives and additives result from the distinct ways in which they interact with contextual cues and discourse in general. For this purpose, I will discuss the claims in light of (a simplified version of) Abrusán (2011, 2016)’s account – an account of presuppositions which is based on the assumption that what content ends up being presupposed is whatever interlocutors do not pay attention to.

According to Abrusán (2011, 2016), interlocutors’ usually pay attention to the main point(s) of a given utterance. There are two ways in which main points can be introduced: (i) the default main point is determined by grammatical constraints (bottom-up), and (ii) the pragmatic secondary main point by contextual cues (top-down). Content that is part of either of the two main points can be understood as being relevant to the overall discourse question and at-issue. In contrast, content that the interlocutors do not pay attention to ends up being presupposed.

Abrusán (2016) argues that, although the presuppositions of both factive verbs and additive particles are triggered by the same grammatical mechanism, they differ from each other with regard to the way they interact with contextual cues. By default and in the absence of extra contextual information, the complements of factive verbs are presupposed, since the primary main point of an utterance is not about the content of the complement. However, interlocutors’ attention can be shifted towards presupposed content by contextual factors, e.g. focus¹, which makes the presupposed content a secondary main point and thereby renders it at-issue. This has been argued to explain the suspension of factives in entailment cancelling environments. For illustration of this process see (83), taken from Beaver (2010) and analysed by Abrusán (2011).

(83) a. If the TA discovers that [your work is plagiarised]ₕ, I will be [forced to notify the Dean]ₕ.
   b. If the TA [discoversₕ] that your work is plagiarised, I will be [forced to notify the Dean]ₕ.

According to Abrusán (2011) the default main point in both (83a–83b) is about the TA discovering that X, where X stands for the embedded clause. In (83a), the embedded clause *your work is plagiarised* is focused and introduces a secondary main point by addressing the background question ‘What will the TA discover?’. Since the interlocutors’ attention is therefore shifted towards the content of the embedded clause *your work is plagiarised* is not presupposed. This contrasts with (83b) where focus indicates the background question ‘What will the TA do?’, and thus corresponds to

¹I understand focus in the sense of Jackendoff (1972), as a theoretical notion that describes the relation between prosodic prominence and pragmatic/semantic effects. Pragmatically, focused material generally addresses the explicit or implicit Question under Discussion (qud). Following Roberts (1996), the QUD corresponds to the current discourse topic.
the default main point. Here, the embedded clause *your work is plagiarised* is being presupposed. Thus, the status of content as being presupposed is malleable to discourse factors such as focus.

Part I hypothesised that both the presuppositional status and the at-issueness of content determine what content is being picked up in the subsequent discourse. This aligns with Abrusán (2011, 2016)’s account: hearers may not pay attention to presupposed content by default and, if applicable, accommodate presupposed content by default. However, as soon as the hearers’ attention is shifted towards that content (because it was made at-issue), it is rendered available for further discussion.

Part II tested whether speakers would exploit the tendency of presupposed content to not usually be the main point. The observed scenario effects suggest that speakers may use the factive verb *know* to sneak in controversial content into the conversation. Based on Abrusán (2016)’s account, one could assume that the outcome of such a strategy relies on the possibility that hearers will not direct their attention towards the controversial and presupposed material. The ease with which the complements of factives can be focused, together with the high objection rate results for presupposed at-issue content (Part I), suggests that the success rate for this strategy is low. This may be yet another reason – besides potentially following higher-order strategies such as coming across as reliable throughout the interaction – for the inter-subject variability in the production data of Part II: speakers may only be occasionally willing to rely on a rather inconsistent strategy.

Example (83) shows that the complements of factives like *know* can be brought to the focus of attention fairly easily. Additive particles are known to interact with focus as well, but in a different way: the focus structure of the host sentence can change what ends up being presupposed. For example, in (84), the focus is on ‘Donald’ and the presupposition arises that someone else other than Donald ate mushrooms which will be satisfied by ‘Mattia’ in the preceding utterance. In contrast, in (85), ‘beetroot’ is focused and that Mattia ate something else other than beetroot is presupposed.

(83) a. Speaker1: Mattia ate mushrooms. Who else ate mushrooms?
   b. Speaker2: [Donald]F ate mushrooms, too.

   *presupposes: There is someone else other than Donald who ate mushrooms*

(85) a. Speaker1: Mattia ate mushrooms. What else did he eat?

---

2This analysis and what is considered to be the evoked background question is based on Abrusán (2011). Alternatively one may argue that the focus on the matrix verb evokes alternatives on the attitudinal/evidential relationship between subject and embedded proposition indicating a background question such as *What is the TA’s attitude/belief about that your work is plagiarised?*. Such a background question still corresponds to the default main point and the interlocutor’s attention is not shifted towards the embedded proposition, questioning whether the work is plagiarised or not.
b. Speaker2: Mattia ate \([\text{beetroot}]_F\), too.

*presupposes:* Mattia ate something else other than beetroot

Abrusán (2016) argues that since the presupposition of additives is generated based on the focus structure of the host sentence, a change in the focus structure would only change the presupposition itself. Moreover, in contrast to factive verbs, presuppositions triggered by additives are not expressed by a constituent and therefore impossible to focus which means that the presupposed content will never become a secondary main point. Consequently, the presupposition triggered by additives cannot be suspended in entailment cancelling environments. Thus, one could argue that the anaphoricity of additive presuppositions prevents the presupposed content from being directly relevant to the discourse topic.

Instead, the production and comprehension of additives is influenced and tied to what has previously been mentioned and is part of the common ground. Specifically, (i) the felicitous production of additives depends on whether there is a suitable antecedent in the preceding context/common ground, which is why presupposed content triggered by additives cannot be accommodated when uttered out of the blue; (ii) the production of additives and the resolution of their anaphoric component was found to be sensitive to content in the common ground (the antecedent), e.g. the extent to which the antecedent is salient. The results of Part III suggest that the antecedent’s salience is partly determined by Similarity and Turn Distance where these notions are applicable.

All in all, a model that aims to capture the production and comprehension of factives and additives should take into account the ways in which these two triggers interact differently with contextual cues and the discourse. More specifically, the comprehension of presuppositions triggered by factives depends on the discourse topic and on what content is rendered at-issue. Likewise, on the production side, factives may be used to structure the subsequent discussion, potentially influencing the upcoming discourse topic (which make them discourse structuring tools). In contrast, while what gets presupposed by additives is influenced by focus placement, the presupposed content cannot be part of the main point or being made at-issue. However, both the production and comprehension of additives is influenced by what content is already part of the common ground. Due to their anaphoric component, additives are used to maintain or manage what has been said rather than structure what interlocutors discuss next (which make them discourse managing tools). However, whether additive presupposition triggers are or are not used to influence what content is discussed next remains to be investigated.
6.2 Discourse structuring versus managing

This chapter discusses the functions of factives and additives being either discourse structuring or managing tools within Farkas and Bruce (2010)’s theory on communication. Farkas and Bruce (2010) introduce a model in form of a table which records what is at-issue in the conversation. The items on the table form a stack, and the goal of the interlocutors is to reduce this stack and increase the common ground (which includes the set of proposition that have been agreed on by all participants and the propositions representing the shared background knowledge of the interlocutors). A speaker who asserts content puts an item on the stack and publicly commits to the asserted proposition. Farkas and Bruce (2010) assumes that an assertion projects confirmation (i.e. is biased towards confirmation) and thereby a future common ground that includes the asserted proposition. Thus, the default of asserting content is the acceptance of the asserted propositions which does not necessarily have to be signalled acceptance. When accepting a proposition, all interlocutors mutually commit to that proposition which is then removed from the stack and enters the common ground. When having entered the common ground, the proposition is removed from the individual commitment lists. Thus, for presupposition satisfaction it is the common ground that counts and not the interlocutors’ commitment lists. In case interlocutors do not jointly commit to a proposition that a speaker asserted, the proposition does not enter the common ground. If interlocutors have agreed to disagree, the proposition is removed from the item stack and only part of the public beliefs of the speaker and those who agree with the speaker.

Farkas and Bruce (2010) do not explicitly discuss presuppositions but I will aim to extent their model to what has been found in this thesis. Our results suggest that speakers’ motives to presuppose rather than assert content depends on their communicative goals. Speakers who aim to be cooperative and informative, should only presuppose content that is already part of the common ground or uncontroversial. New information that speakers presuppose pose a challenge for the evaluation of the whole utterance and its proposition(s): the presupposition has to be part of the common ground in order for the whole utterance to be evaluated. Assuming for now the presupposed content is not at-issue it should not be part of the at-issue stack to begin with. One could then argue – since confirmation is the default for assertions – that the default for presuppositions is accommodation. Accommodation would then involve simply adding the presupposed content to the common ground after which the at-issue content can be evaluated. The difference between accommodation of presuppositions and confirmation of assertions would be that accommodation is always implicit, whereas confirmation may be implicit or signalled (e.g. by ‘yes’, ‘I agree’ etc).

In contrast, new and potentially controversial content should be asserted and thereby signalled to
be available for further discussion. However, our results suggest that when speakers’ communicative goals change and their behaviour is no longer fully cooperative, speakers more frequently exploit the content-backgrounding faculties of (factive) presupposition triggers to sneak potentially controversial content into the conversation. With respect to Farkas and Bruce (2010)’s model, one could argue that in such a case speakers rely on the accommodation default or the bias to accommodate presupposed content. If such a strategy was successful, speakers successfully diverted adding the presupposed content to the stack where it had to be discussed. Instead, the presupposed content was added to the common ground. In this sense speakers used presupposing to structure the subsequent discourse. However, this strategy is far from being bullet-proof, since in case the presupposed content is perceived to be at-issue (which controversial content most likely is) by the interlocutors, the presupposition should be part of the at-issue stack. When added to the at-issue stack, interlocutors should treat the presupposed proposition similar to asserted propositions on the stack and either confirm (implicitly or explicitly) or deny (explicitly) the proposition.

Similarly, one could understand presupposing as speakers’ cooperative attempt to engage in grounding by referring back to content of the common ground. Especially additive presupposition triggers seem to play an essential role in the interlocutors’ grounding process due to their anaphoric component. Applied to Farkas and Bruce (2010)’s model, interlocutors have to keep track of what content is part of the common ground in order to embed a given proposition in a context. More specifically, grounding helps to re-iterate what content individual interlocutors have publicly committed to, and what content all interlocutors have jointly committed to. Similarly, grounding may serve the speaker and others to keep private beliefs apart from mutually shared beliefs. Our findings that interlocutors did not always engage in grounding via additive usage, may suggest that sometimes there is a dissociation between interlocutors’ private beliefs and mutually shared beliefs.

6.3 Methodology

The experiments of this thesis were mostly designed to create situations that were as close as possible to actual interaction, which was arguably easier to accomplish for in-person testing (Part I) than for online testing (Parts II and III). In this chapter, I will review a few open questions, aspects and shortcomings of the methodology used in this thesis.

In Part I, a dialogue-based experiment [1] was developed which was carried out in-person. Recall that participants were asked to take over the role of a detective in a fictional interrogation whose task is to interrogate a suspect (confederate). Participants were instructed to (i) ask the suspect scripted questions as well as (ii) identify and challenge lies in the suspect’s responses based on pieces
of evidence. The role of the suspect was played by a confederate (the same person throughout the experiment) who was informed about the purpose of the experiment and was familiar with the experimental material and the scripted answers. Despite instructing the confederate to answer the participants’ question in the same way throughout the experiment, it is likely that there was some degree of variation, e.g. in placement of stress. As was mentioned before, focus placed via prosodic prominence may introduce a secondary main point (Abruşán, 2011, 2016) and render content at-issue. For example, in (86b) as opposed to (86a) the focus on ‘Russia’ might evoke the background question about where the suspect stopped working. This makes the presupposed content that the suspect used to work in Russia potentially more at-issue than for (86a). In stop, with a focus on stop the background question would be whether the suspect stopped working for the national gallery in Russia, which does not render the presupposed content at-issue. Thus, inconsistent stress placement might have made presupposed content more or less available for discussion. A way to control for variation in prosody would have been to use recordings. However, by letting participants interact with an actual person, a more naturalistic and interactive experiment was created which elicited comparatively natural and spontaneous speech.

(86) Have you held any other positions?

The potential importance of focus similarly applies to the online experiments [2–5] of Part II. For all experiments, participants were presented with written utterances to either choose from (production) or to rate for speaker confidence (comprehension). This means that participants most likely placed their focus individually which might have introduced some variation. For example, focusing different constituents introduces implicit background questions rendering different parts of the sentence at-issue, see (87): the focus on suspect gives rise to the question ‘Who deposited money?’ (87b), and the focus on Edinburgh would give rise to the question ‘At what station did the suspect deposit money?’ (87d).

(87) Briefing item:
   Participant: What happened on Saturday the 30th of November 2019?
   Colleague: The suspect went to visit friends in Glasgow over the weekend.
   Participant:
      a. I [know]₇ that the suspect deposited money in one of the lockers at Edinburgh station.
      b. I know that the [suspect]₇ deposited money in one of the lockers at Edinburgh station.
      c. I know that the suspect deposited [money]₇ in one of the lockers at Edinburgh station.
      d. I know that the suspect deposited money in one of the lockers at [Edinburgh]₇ station.
Rendering different parts of the presupposed content conveyed by the complement clause at-issue, see examples (87b–87d), seems to introduce some uncertainty despite of know being associated with high degrees of belief. For example, the focus on ‘money’ in (87c) casts doubt on what exactly the suspect deposited at the station, assuming that such a focus placement indicates the background question What did the subject deposit in the locker? and evokes focus alternatives, for example, the suspect could have deposited luggage, the painting, a note for the accomplice etc. instead of money. However, I can only speculate in what way individual focus placement in experiments of Part II might have resulted in between- and perhaps within-subject variation. To my knowledge, studies investigating presuppositions and focus placement have so far concentrated on presupposition projection, and thereby, on presuppositions in entailment cancelling environments (e.g. Cummins & Rohde, 2015; Tonhauser, 2016). Applied to the production experiments [2–4], it remains to be investigated whether speakers would utter (87a) with a focus on the factive presupposition trigger for higher degrees of belief than utterances (87b–87d). Applied to the comprehension experiment [5], it remains to be investigated whether hearers understand a speaker uttering (87a) to be more certain than a speaker uttering (87b–87d).

As was mentioned earlier, what ends up being presupposed by additives is influenced by the placement of focus (e.g. Rooth, 1985). For example, in (88a) too associates with the prosodically stressed ‘I’ which makes the presupposed propositional alternative about an alternative subject. In contrast, in (88b), too associates with ‘Lisa’ which makes the propositional alternative about who is starting in January.

(88) a. [I]F remember that Lisa will move to Konstanz, too.
   presupposes: Someone other than the speaker remember that Lisa will move to Konstanz.

b. I remember that [Lisa]F’s will move to Konstanz, too.
   presupposes: The speaker remembers that someone other than Lisa will move to Konstanz.

As was previously discussed in Chapter 5, since participants were not provided with audio recordings in experiments [6–7], they had to assign their focus based on their own judgement, e.g. participant1 may have placed their focus on ‘I’ (88a) and participant2 on ‘Lisa’ (88b). Triggering different presuppositions may have led to different assessments by participants of whether there was a suitable antecedent in the preceding discourse, which would in turn induce some variation in additive usage between participants. It remains to be investigated whether and how focus placement actually influences the production of additives in a setting such as the one in experiments [6–7].

Overall, the studies of this thesis attempted to create intuitive and/or naturalistic situations in
order to elicit responses that were as close to natural speech as possible. I think that it is crucial to develop engaging experimental designs in order to capture the subtle effects that context has on interaction. This is even more crucial for online experiments were participants often take part in multiple experiments per day which means that efforts have to be made to keep their attention. I hope that the material developed and discussed in this thesis, which has been made available online, will usefully inform future experimental designs.

### 6.4 Relevance for other areas

This thesis investigated the production and comprehension of presuppositions and assertions by taking into account a wide range of communicative goals such as deceiving, convincing, informing or hedging. The results of this thesis help us to better understand how speakers can utilise presuppositions when they follow communicative goals other than being informative. In this section I will review the relevance of investigating strategic communication in general and for other linguistic areas.

The experiments of Parts I and II focused on the contrast between cooperative and uncooperative communication and the potential strategic advantage presupposing may have over asserting. I take ‘uncooperative communication’ to represent situations in which speakers try to induce their hearers to believe things that are not true in the world or to prevent their hearers from knowing the truth. In this way, the observed results contribute to studies on speakers’ exploitation of pragmatic phenomena in general and the extent to which speakers can do this to advance uncooperative agendas. More specifically, in agreement with previous research (e.g. Franke et al., 2020) it seems that speakers who strategically exploit the features of presuppositions or contextually inferred meaning such as implicatures still have to follow cooperative principles in order to reach their uncooperative goals successfully. For example, speakers who convey their deception/lies via implicature are cooperative in the narrow sense that they mean this implicature to be recoverable and rely on hearers to be cooperative to the extent that they infer the implicature. Indeed, Franke et al. (2020) found that speakers expect hearers to infer their implicatures even when it is obvious that the speaker is uncooperative. Regarding the uncooperative usage of presuppositions, the findings suggest that while speakers are uncooperative by sneaking controversial content into the hearers’ common ground via presupposition, they similarly rely on the cooperativity of their hearers to accommodate the presupposed content.

Strategic considerations are especially relevant for numerical communication which is essential for any kind of risk assessment usually involving decisions where much is at stake, e.g. discussing
treatments in doctor-patient conversations, deciding on financial investments, or assessing natural
disasters. In cooperative situations, one would assume that speakers aim to reduce the hearer’s
uncertainty. For example, in an ideal world doctors inform patients thoroughly about the risk of
certain treatments etc. However, there are usually additional factors at play. Doctors can be held
accountable for what they say and may be bound to certain government guidelines or contracts with
pharmaceutical companies. So in doctor-patient communication there may be competing goals at
play: informing, adhering to/ avoiding accountability, and convincing.

The interplay between different communicative goals may be illustrated by contrasting the following
utterances that yet again seem very similar on the surface. An utterance such as (89a) may lead
to the inference that the speaker does not have a definite amount in mind: *at least* was argued to
put a range rather than a definite amount under discussion (Nouwen, 2010) which makes stating
a precise percentage such as ‘90% to be precise’ sound odd (highlighted with #). Thus, uttering
(89a) may be useful for doctors to keep their accountability low, since uttering *at least* induces the
impression that the speaker does not have a specific amount in mind which prevents them from
committing to such.

(89) a. The AstraZeneca vaccine is at least 80% effective.
   #90% to be precise.

b. The AstraZeneca vaccine is more than 80% effective.
   90% to be precise.

In contrast, *more than* in (89b) is a comparative quantifier, and compares a precise amount (80%)
to some contrastive amount (90%). Thus, using *more than* places a definite amount under dis-
cussion (Nouwen, 2010) and the continuation ‘90% to be precise’ is felicitous. Hence, a doctor
uttering *more than* may signal that they are able and willing to discuss the exact effectiveness
of the vaccine further. The contrast between these two quantifiers illustrates that speakers may
choose different linguistic means to induce slightly different readings which in turn influence the
subsequent discourse and potentially the hearer’s decision in different ways. However, it remains
to be investigated whether speakers indeed produce quantifiers such as *at least* more frequently to
avoid accountability and committing to a concrete numerical amount, and whether speakers who
want to discuss a concrete amount would produce *more than* more frequently.

Furthermore, Parts II and III investigated how speakers act politely by hedging their statements,
and whether speakers deviate from their cooperative efforts to manage the discourse when faced with
an impolite interlocutor. In Part II we have seen that speakers tend to hedge their certainty when
in a cooperative situation with an interlocutor in a similar power position. However, it is less widely
studied to what extent high-power speakers engage in hedging. Unequal power relations between interlocutors are especially frequent in risk communication, e.g. a doctor is in a high-power position relative to the patient. As was previously mentioned, doctors can be held accountable for what they say, and furthermore, may try to avoid communicating bad news directly in an effort to be compassionate. Thus, I would assume that even high-power speakers engage in hedging. Bonnefon and Villejoubert (2006)’s findings align with this intuition in showing that patients (low-power hearers) appear to be aware that doctors (high-power speaker) may hedge their statements. Yet, it remains to be investigated whether high-power speakers indeed use hedging in risk-communication.

Part III of this thesis investigated the potential social level of additive usage. More specifically, it was tested whether speakers drop the usage of additives when the antecedent speaker deviates from social norms by behaving impolitely. It was of interest to see whether additive production can be seen as a convergence move with the antecedent speaker. Besides affecting additive production, social accommodation and with that the convergence and divergence from speakers may be relevant for numerical communication. In the doctor-patient context, patients who have difficulties with understanding probabilities could potentially see doctors communicating risk by solely using probabilities as signalling divergence. From a doctor’s perspective, diverging from the patient could be understood as a move where the doctor establishes themselves as knowledgeable and situates themselves in a more powerful position relative to the patient that it would be best to rely on the doctor’s judgement.

Moreover, in Part III, we have seen that participants’ production of additive particles deviates to some extent from what has been predicted in theory: (i) speakers uttered additives not as frequently as predicted by accounts postulating obligatoriness, and (ii) speakers use additives even when the antecedent was uttered a while back. One may argue that this deviation simply stems from the fact that the additive rate was measured in an experimental environment instead of actual interaction in the real world and/or that participants did not pay enough attention to whether additives should be used or not. With regard to the first objection, the experiment was, to my knowledge, the first attempt to study production of additives in interaction (beyond text production). With regard to the latter objection, the attention checks as well as the context items suggest that participants paid attention to the conversation throughout the experiment. In general, I think that studying production of other semantic and pragmatic phenomena experimentally (in addition to corpus studies) is essential to inform theoretical accounts. Investigating comprehension primarily through the lens of acceptability or felicity rating studies seems not suffice to understand the whole phenomenon.
Chapter 7

Conclusion

This thesis aimed to discover possible reasons why speakers would choose to presuppose rather than assert content. Results of Part I tentatively indicated that what content is being picked up by interlocutors depends partly on the content’s presuppositional status (as well as its at-issueness). More specifically, hearers accommodated (or accepted) potentially controversial at-issue content more frequently when it was presupposed as opposed to asserted. These results led to the hypothesis, which was further investigated in Part II, that uncooperative speakers may exploit the hearer’s tendency to accommodate presupposed content more readily than accepting asserted content. The findings of Part II indicated that speakers indeed make choices to presuppose a given proposition $p$ (via the formulation $know\ p$) rather than assert $p$ when aiming to maximise the chances of the hearer to accept $p$ without further discussion. Hence, our results align with the hypothesis that factive presupposition triggers can be used as a discourse structuring tool. Part III investigated how and whether the production of additive presupposition triggers such as too is influenced by discourse factors and the speaker’s communicative goals. The results align with the hypothesis that speakers use additives to engage in discourse management and thereby actively participate in the interlocutors’ grounding process.

All in all, whereas presupposing is often a by-product of using language, speakers may favour linguistic formulations (presuppositions over assertions) that allow them to package information in strategic ways and serve their broader discourse goals; either to structure the discourse to their liking (factive) as part of an uncooperative discourse move, or to make an active effort to engage in the grounding process (additive) as part of a cooperative discourse move. To conclude, the research questions of this thesis (RQ, A–D) are part of a larger effort both here and beyond to understand what guides speaker choices in discourse and to test how aware hearers are of these choices.
References


Farkas, D. F. & Bruce, K. B. (2010). On reacting to assertions and polar questions. *Journal of
Semantics, 27(1), 81–118.


Fraser, B. (2010). Pragmatic competence: The case of hedging. In G. Kaltenböck, W. Mihatsch & S. Schneider (Eds.), New approaches to hedging (pp. 15–34). Emerald.


auch (=’too’). *Frontiers in Communication*, 15.


Potts, C. (2008). *Wait a minute! What kind of discourse strategy is this?* (Annotated data)


Appendix
## Appendix A

### 7.1 Experiment [1]: Sample experimental items

<table>
<thead>
<tr>
<th>Trigger</th>
<th>Question</th>
<th>Presupposed content condition</th>
<th>Asserted content condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>stop</td>
<td>Have you held any other positions?</td>
<td>I stopped working for the national gallery in Russia until two thousand and seventeen.</td>
<td>I used to work for the national gallery in Russia until two thousand and seventeen.</td>
</tr>
<tr>
<td>know</td>
<td>When did John Smith first approach you in two thousand eighteen?</td>
<td>I know that I first saw him at a gallery opening in November.</td>
<td>I first saw him at a gallery opening in November.</td>
</tr>
<tr>
<td>to be happy</td>
<td>We are interested in what happened on the 2nd of September. What did you do on that day?</td>
<td>I was happy that I had time to finish a journal article at home.</td>
<td>I had time to finish a journal article at home.</td>
</tr>
<tr>
<td>discover</td>
<td>Now, we would like to know more about the forth of September when you went to the locksmith. What happened there?</td>
<td>I discovered that the key to my flat needed replacing.</td>
<td>The key to my flat needed replacing.</td>
</tr>
<tr>
<td>regret</td>
<td>How was the race on the seventeenth of October?</td>
<td>I regret that I didn’t go to that race.</td>
<td>I didn’t go to that race.</td>
</tr>
<tr>
<td>to be annoyed</td>
<td>Did you meet John Smith on that day?</td>
<td>I was annoyed that I bumped into him unexpectedly.</td>
<td>I bumped into him unexpectedly.</td>
</tr>
<tr>
<td>only</td>
<td>Did you stay at the race-course after going to the bar?</td>
<td>I only went to the bathroom at four pm.</td>
<td>I went to the bathroom at four pm.</td>
</tr>
<tr>
<td>return</td>
<td>We would like to know more about the seventeenth of October. What else did you do that day?</td>
<td>I returned to my office.</td>
<td>Before and after the race I was at the office.</td>
</tr>
</tbody>
</table>

Table 7.1: Experimental items for experiment [1].
Appendix B

7.2 Experiments [2–3]: 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 7.1: Briefing item for experiments [2–3] (suspect Emily Brown, weak evidence).
Did Emily Brown have any financial problems?

Financially the suspect was doing alright.

Evidence:

HARRIOT BANK

Statement Ending 30/11/2019

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

Summary of Accounts

<table>
<thead>
<tr>
<th>Account Type</th>
<th>Account Number</th>
<th>Ending Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td>HARRIOT RELATIONSHIP CHECKING</td>
<td>XXXXXXXX0000151</td>
<td>£ -30,040</td>
</tr>
</tbody>
</table>

Primary Checking

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>01/01/2020</td>
<td>Beginning Balance</td>
<td>£ 1,000</td>
</tr>
<tr>
<td>10/12/2020</td>
<td>Ending Balance</td>
<td>£ -25,005</td>
</tr>
</tbody>
</table>

Figure 7.2: Briefing item for experiments [2–3] (suspect Emily Brown, strong evidence).
Did you have any financial problems?

Financially I was doing alright.

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 7.3: Interrogation item for experiments [2–3] (suspect Emily Brown, weak evidence).
Did you have any financial problems?

Financially I was doing alright.

Evidence:

Emily Brown’s bank statement from November 2019

Figure 7.4: Interrogation item for experiments [2–3] (suspect Emily Brown, strong evidence).
Where was Johanna Smith on the night of the theft?

She was at a birthday party the whole night.

Evidence:

None of the party guests can recall whether Johanna S. stayed until the end.

-Campbell

Figure 7.5: Briefing item for experiments [2-3] (suspect Johanna Smith, weak evidence).
Where was Johanna Smith on the night of the theft?

She was at a birthday party the whole night.

Evidence:

Video footage of Johanna Smith heading towards the national gallery during the time of the party.

Figure 7.6: Briefing item for experiments [2–3] (suspect Johanna Smith, strong evidence).
Where were you on the night of the theft?

I was at a birthday party the whole night.

Evidence:

None of the party guests can recall whether Johanna S. stayed until the end.

— Campbell

Figure 7.7: Interrogation item for experiments [2–3] (suspect Johanna Smith, weak evidence).
Where were you on the night of the theft?

I was at a birthday party the whole night.

Evidence:

Video footage of Johanna Smith heading towards the national gallery during the time of the party.

Figure 7.8: Interrogation item for experiments [2–3] (suspect Johanna Smith, strong evidence).

7.3 Experiments [2–3]: Additional Analyses
<table>
<thead>
<tr>
<th>Utterance</th>
<th>Coefficient</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>$\hat{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>notice</td>
<td>Intercept</td>
<td>-1.39</td>
<td>0.31</td>
<td>-2.04</td>
<td>-0.84</td>
<td>1.00</td>
</tr>
<tr>
<td>am sure</td>
<td>Intercept</td>
<td>-2.02</td>
<td>0.45</td>
<td>-2.99</td>
<td>-1.23</td>
<td>1.00</td>
</tr>
<tr>
<td>see</td>
<td>Intercept</td>
<td>-0.96</td>
<td>0.24</td>
<td>-1.48</td>
<td>-0.53</td>
<td>1.00</td>
</tr>
<tr>
<td>think</td>
<td>Intercept</td>
<td>-0.68</td>
<td>0.25</td>
<td>-1.19</td>
<td>-0.20</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Intercept</td>
<td>0.05</td>
<td>0.22</td>
<td>-0.39</td>
<td>0.46</td>
<td>1.00</td>
</tr>
<tr>
<td>guess</td>
<td>Intercept</td>
<td>-6.84</td>
<td>1.88</td>
<td>-11.35</td>
<td>-4.05</td>
<td>1.00</td>
</tr>
<tr>
<td>other</td>
<td>Intercept</td>
<td>-3.87</td>
<td>0.84</td>
<td>-5.80</td>
<td>-2.49</td>
<td>1.00</td>
</tr>
<tr>
<td>notice</td>
<td>Evidentiality</td>
<td>-0.38</td>
<td>0.27</td>
<td>-0.89</td>
<td>0.18</td>
<td>1.00</td>
</tr>
<tr>
<td>am sure</td>
<td>Evidentiality</td>
<td>-0.61</td>
<td>0.26</td>
<td>-1.11</td>
<td>-0.09</td>
<td>1.00</td>
</tr>
<tr>
<td>see</td>
<td>Evidentiality</td>
<td>0.01</td>
<td>0.22</td>
<td>-0.40</td>
<td>0.46</td>
<td>1.00</td>
</tr>
<tr>
<td>think</td>
<td>Evidentiality</td>
<td>-1.11</td>
<td>0.19</td>
<td>-1.50</td>
<td>-0.74</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Evidentiality</td>
<td>-0.75</td>
<td>0.19</td>
<td>-1.13</td>
<td>-0.36</td>
<td>1.00</td>
</tr>
<tr>
<td>guess</td>
<td>Evidentiality</td>
<td>-1.02</td>
<td>0.62</td>
<td>-2.18</td>
<td>0.32</td>
<td>1.00</td>
</tr>
<tr>
<td>other</td>
<td>Evidentiality</td>
<td>-0.78</td>
<td>0.42</td>
<td>-1.54</td>
<td>0.09</td>
<td>1.00</td>
</tr>
<tr>
<td>notice</td>
<td>Scenario1</td>
<td>0.17</td>
<td>0.19</td>
<td>-0.21</td>
<td>0.54</td>
<td>1.00</td>
</tr>
<tr>
<td>am sure</td>
<td>Scenario1</td>
<td>0.82</td>
<td>0.26</td>
<td>0.35</td>
<td>1.38</td>
<td>1.00</td>
</tr>
<tr>
<td>see</td>
<td>Scenario1</td>
<td>0.11</td>
<td>0.16</td>
<td>-0.22</td>
<td>0.43</td>
<td>1.00</td>
</tr>
<tr>
<td>think</td>
<td>Scenario1</td>
<td>0.65</td>
<td>0.17</td>
<td>0.33</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Scenario1</td>
<td>0.44</td>
<td>0.14</td>
<td>0.16</td>
<td>0.72</td>
<td>1.00</td>
</tr>
<tr>
<td>guess</td>
<td>Scenario1</td>
<td>1.09</td>
<td>0.57</td>
<td>-0.05</td>
<td>2.25</td>
<td>1.00</td>
</tr>
<tr>
<td>other</td>
<td>Scenario1</td>
<td>0.73</td>
<td>0.46</td>
<td>-0.19</td>
<td>1.67</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 7.2: Experiment 1 [2]: Population-level estimates of the categorical regression model including responses *guess* and *other* 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.
<table>
<thead>
<tr>
<th>Utterance</th>
<th>Coefficient</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>$\hat{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>believe</td>
<td>Intercept</td>
<td>0.41</td>
<td>0.17</td>
<td>0.07</td>
<td>0.75</td>
<td>1.00</td>
</tr>
<tr>
<td>other</td>
<td>Intercept</td>
<td>-2.90</td>
<td>0.38</td>
<td>-3.75</td>
<td>-2.20</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Evidentiality</td>
<td>-1.21</td>
<td>0.14</td>
<td>-1.49</td>
<td>-0.93</td>
<td>1.00</td>
</tr>
<tr>
<td>other</td>
<td>Evidentiality</td>
<td>-0.91</td>
<td>0.22</td>
<td>-1.31</td>
<td>-0.46</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Scenario1</td>
<td>0.25</td>
<td>0.09</td>
<td>0.08</td>
<td>0.42</td>
<td>1.00</td>
</tr>
<tr>
<td>other</td>
<td>Scenario1</td>
<td>0.24</td>
<td>0.24</td>
<td>-0.24</td>
<td>0.71</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 7.3: Experiment 2 [3]: Population-level estimates of the categorical regression model including responses *guess* and *other* 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.
Appendix C

7.4 Production experiment [4]

7.4.1 Production experiment [4]: Model output with uninformative priors

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

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

Table 7.4: Production experiment [experiment 4]: 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.
### Production experiment [4]: Model output with order effect

Table 7.5: Production experiment [experiment 4]: 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.

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Coefficient</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>$\hat{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>believe</td>
<td>Intercept</td>
<td>1.18</td>
<td>0.26</td>
<td>0.68</td>
<td>1.70</td>
<td>1.00</td>
</tr>
<tr>
<td>bare assertion</td>
<td>Intercept</td>
<td>0.30</td>
<td>0.23</td>
<td>-0.15</td>
<td>0.73</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Evidentiality</td>
<td>-1.51</td>
<td>0.17</td>
<td>-1.86</td>
<td>-1.19</td>
<td>1.00</td>
</tr>
<tr>
<td>bare assertion</td>
<td>Evidentiality</td>
<td>-0.36</td>
<td>0.16</td>
<td>-0.67</td>
<td>-0.04</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Scenario</td>
<td>0.59</td>
<td>0.11</td>
<td>0.39</td>
<td>0.80</td>
<td>1.00</td>
</tr>
<tr>
<td>bare assertion</td>
<td>Scenario</td>
<td>0.27</td>
<td>0.11</td>
<td>0.06</td>
<td>0.49</td>
<td>1.00</td>
</tr>
<tr>
<td>believe</td>
<td>Order</td>
<td>0.00</td>
<td>0.03</td>
<td>-0.05</td>
<td>0.06</td>
<td>1.00</td>
</tr>
<tr>
<td>bare assertion</td>
<td>Order</td>
<td>-0.01</td>
<td>0.03</td>
<td>-0.06</td>
<td>0.05</td>
<td>1.00</td>
</tr>
</tbody>
</table>
7.4.3 Production experiment [4]: Model predictions with interaction

![Graph](image_url)

(a) Facet Utterance

(b) Facet Scenario

Figure 7.9: Production experiment [experiment 4]: Predictions of the interaction model given the data. 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. The lines represents the means of the fixed effects and the shaded area depicts the 95% credible interval of the fixed effects.
7.4.4 Production experiment [4]: 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 7.10: Briefing item for the production experiment [experiment 4] (suspect Emily Brown, weak evidence).
Figure 7.11: Briefing item for the production experiment [experiment 4] (suspect Emily Brown, strong evidence).
Did you have any financial problems?

Financially I was doing alright.

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 7.12: Interrogation item for the production experiment [experiment 4] (suspect Emily Brown, weak evidence).
Figure 7.13: Interrogation item for the production experiment [experiment 4] (suspect Emily Brown, strong evidence).

7.5 Comprehension experiment [5]

7.5.1 Comprehension experiment [5]: 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)
<table>
<thead>
<tr>
<th>Coefficient</th>
<th>posterior mean</th>
<th>Standard Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>$\hat{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (grand mean)</td>
<td>1.06</td>
<td>0.10</td>
<td>0.87</td>
<td>1.24</td>
<td>1.00</td>
</tr>
<tr>
<td>Utterance I</td>
<td>-0.83</td>
<td>0.08</td>
<td>-0.99</td>
<td>-0.68</td>
<td>1.00</td>
</tr>
<tr>
<td>Utterance II</td>
<td>0.17</td>
<td>0.06</td>
<td>0.06</td>
<td>0.28</td>
<td>1.00</td>
</tr>
<tr>
<td>ScenarioI</td>
<td>-0.07</td>
<td>0.04</td>
<td>-0.16</td>
<td>0.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 7.6: Comprehension experiment [experiment 5]: 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).

7.5.2 Comprehension experiment [5]: Model output with order effect

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>posterior mean</th>
<th>Standard Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>$\hat{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (grand mean)</td>
<td>1.07</td>
<td>0.10</td>
<td>0.86</td>
<td>1.27</td>
<td>1.00</td>
</tr>
<tr>
<td>Utterance I</td>
<td>-0.83</td>
<td>0.08</td>
<td>-0.98</td>
<td>-0.68</td>
<td>1.00</td>
</tr>
<tr>
<td>Utterance II</td>
<td>0.16</td>
<td>0.06</td>
<td>0.05</td>
<td>0.28</td>
<td>1.00</td>
</tr>
<tr>
<td>ScenarioI</td>
<td>-0.07</td>
<td>0.04</td>
<td>-0.16</td>
<td>0.01</td>
<td>1.00</td>
</tr>
<tr>
<td>Order</td>
<td>-0.00</td>
<td>0.01</td>
<td>-0.02</td>
<td>0.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 7.7: Comprehension experiment [experiment 5]: 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 (4.6).
7.5.3 Comprehension experiment [5]: Sample experimental items

Did Emily Brown have any financial problems?

Financially the suspect was doing alright.

Colleague 1

I know that the suspect was in need of money.

Colleague 2

Figure 7.14: Briefing item for the comprehension experiment [5] (suspect Emily Brown/‘know’).
Figure 7.15: Interrogation item for the comprehension experiment [5] (suspect Emily Brown/’believe’).
Figure 7.16: Interrogation item for the comprehension experiment [5] (suspect Johanna Smith/‘know’).
7.5.4 Comprehension experiment [5]: Sample control items

Figure 7.17: Briefing control item/attention check for the comprehension experiment [experiment 5].

Figure 7.18: Interrogation control item/attention check for the comprehension experiment [experiment 5].
Appendix D

7.6 Experiment I [6]

7.6.1 Experiment I [6]: Model output with uninformative priors

- by-expression intercepts: normal(0, 10)
- fixed effects: normal(0, 10)
- standard deviations: normal(0, 10)
- correlation matrix: LKJ(2)

<table>
<thead>
<tr>
<th>Coefficient Estimates</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-9.60</td>
<td>3.02</td>
<td>-16.55</td>
<td>-4.87</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity</td>
<td>6.98</td>
<td>2.93</td>
<td>2.48</td>
<td>13.80</td>
<td>1.00</td>
</tr>
<tr>
<td>Turn Distance</td>
<td>6.25</td>
<td>2.89</td>
<td>1.90</td>
<td>13.02</td>
<td>1.00</td>
</tr>
<tr>
<td>Politeness</td>
<td>0.21</td>
<td>2.84</td>
<td>-5.50</td>
<td>5.97</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity*Turn Distance</td>
<td>-5.31</td>
<td>2.90</td>
<td>-12.03</td>
<td>-0.87</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity*Politeness</td>
<td>0.41</td>
<td>2.84</td>
<td>-5.41</td>
<td>6.13</td>
<td>1.00</td>
</tr>
<tr>
<td>Turn Distance*Politeness</td>
<td>-0.19</td>
<td>2.84</td>
<td>-5.98</td>
<td>5.53</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity<em>Turn Distance</em>Politeness</td>
<td>0.19</td>
<td>2.84</td>
<td>-5.53</td>
<td>5.93</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 7.8: Logistic regression output for experiment I [6] with wider prior for the intercept.
7.6.2 Experiment I [6]: Model output with order effect

<table>
<thead>
<tr>
<th></th>
<th>Block2</th>
<th>Block3</th>
<th>Block4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Block2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Block3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Block4</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

Table 7.9: Sum-coding for the predictor Block order for experiments I and II [6-7].

<table>
<thead>
<tr>
<th>Coefficient Estimates</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>( \hat{R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.92</td>
<td>0.53</td>
<td>-5.05</td>
<td>-2.95</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity</td>
<td>1.59</td>
<td>0.43</td>
<td>0.73</td>
<td>2.47</td>
<td>1.00</td>
</tr>
<tr>
<td>Turn Distance</td>
<td>1.17</td>
<td>0.37</td>
<td>0.51</td>
<td>1.96</td>
<td>1.00</td>
</tr>
<tr>
<td>Politeness</td>
<td>0.17</td>
<td>0.35</td>
<td>-0.52</td>
<td>1.96</td>
<td>1.00</td>
</tr>
<tr>
<td>Block1</td>
<td>0.03</td>
<td>0.49</td>
<td>-0.94</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Block2</td>
<td>0.45</td>
<td>0.49</td>
<td>-0.60</td>
<td>1.36</td>
<td>1.00</td>
</tr>
<tr>
<td>Block3</td>
<td>-0.91</td>
<td>0.54</td>
<td>-1.99</td>
<td>0.17</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity*Turn Distance</td>
<td>-0.42</td>
<td>0.40</td>
<td>-1.24</td>
<td>0.36</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity*Politeness</td>
<td>0.35</td>
<td>0.35</td>
<td>-0.35</td>
<td>1.05</td>
<td>1.00</td>
</tr>
<tr>
<td>Turn Distance*Politeness</td>
<td>-0.09</td>
<td>0.35</td>
<td>-0.81</td>
<td>0.61</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity<em>TurnDistance</em>Politeness</td>
<td>0.10</td>
<td>0.35</td>
<td>-0.58</td>
<td>0.79</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 7.10: Logistic regression output for experiment I [6] with Block order as a predictor.

7.7 Experiment II [7]

7.7.1 Experiment II [7]: Model output with uninformative priors

- by-expression intercepts: normal(0, 10)
- fixed effects: normal(0, 10)
- standard deviations: normal(0, 10)
- correlation matrix: LKJ(2)
<table>
<thead>
<tr>
<th>Coefficient Estimates</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>$\hat{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.54</td>
<td>0.42</td>
<td>-2.43</td>
<td>-0.74</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity</td>
<td>1.59</td>
<td>0.48</td>
<td>0.68</td>
<td>2.60</td>
<td>1.00</td>
</tr>
<tr>
<td>Turn Distance</td>
<td>0.85</td>
<td>0.26</td>
<td>0.39</td>
<td>1.43</td>
<td>1.00</td>
</tr>
<tr>
<td>Politeness</td>
<td>0.43</td>
<td>0.20</td>
<td>0.08</td>
<td>0.85</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity*Turn Distance</td>
<td>-0.22</td>
<td>0.21</td>
<td>-0.65</td>
<td>0.21</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity*Politeness</td>
<td>-0.02</td>
<td>0.17</td>
<td>-0.36</td>
<td>0.31</td>
<td>1.00</td>
</tr>
<tr>
<td>Turn Distance*Politeness</td>
<td>-0.28</td>
<td>0.20</td>
<td>-0.71</td>
<td>0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity<em>Turn Distance</em>Politeness</td>
<td>0.27</td>
<td>0.15</td>
<td>-0.01</td>
<td>0.58</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 7.11: Logistic regression output for experiment II [7] with wider priors I.

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

<table>
<thead>
<tr>
<th>Coefficient Estimates</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>$\hat{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.49</td>
<td>0.40</td>
<td>-2.32</td>
<td>-0.70</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity</td>
<td>1.56</td>
<td>0.47</td>
<td>0.66</td>
<td>2.51</td>
<td>1.00</td>
</tr>
<tr>
<td>Turn Distance</td>
<td>0.85</td>
<td>0.26</td>
<td>0.39</td>
<td>1.41</td>
<td>1.00</td>
</tr>
<tr>
<td>Politeness</td>
<td>0.42</td>
<td>0.20</td>
<td>0.07</td>
<td>0.84</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity*Turn Distance</td>
<td>-0.22</td>
<td>0.22</td>
<td>-0.67</td>
<td>0.22</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity*Politeness</td>
<td>-0.03</td>
<td>0.17</td>
<td>-0.35</td>
<td>0.30</td>
<td>1.00</td>
</tr>
<tr>
<td>Turn Distance*Politeness</td>
<td>-0.28</td>
<td>0.20</td>
<td>-0.69</td>
<td>0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity<em>Turn Distance</em>Politeness</td>
<td>0.27</td>
<td>0.15</td>
<td>-0.01</td>
<td>0.57</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 7.12: Logistic regression output for experiment II [7] with wider priors II.
### 7.7.2 Experiment II [7]: Model output with order effect

<table>
<thead>
<tr>
<th>Coefficient Estimates</th>
<th>posterior mean</th>
<th>Est.Error</th>
<th>l-95% CrI</th>
<th>u-95% CrI</th>
<th>( \hat{R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.45</td>
<td>0.23</td>
<td>-1.90</td>
<td>-1.01</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity</td>
<td>1.29</td>
<td>0.38</td>
<td>0.53</td>
<td>2.02</td>
<td>1.00</td>
</tr>
<tr>
<td>TurnDistance</td>
<td>0.74</td>
<td>0.20</td>
<td>0.34</td>
<td>1.15</td>
<td>1.00</td>
</tr>
<tr>
<td>Politeness</td>
<td>0.36</td>
<td>0.16</td>
<td>0.05</td>
<td>0.71</td>
<td>1.00</td>
</tr>
<tr>
<td>Block1</td>
<td>-0.28</td>
<td>0.33</td>
<td>-0.96</td>
<td>0.35</td>
<td>1.00</td>
</tr>
<tr>
<td>Block2</td>
<td>0.53</td>
<td>0.28</td>
<td>-0.10</td>
<td>1.06</td>
<td>1.00</td>
</tr>
<tr>
<td>Block3</td>
<td>-0.88</td>
<td>0.36</td>
<td>-1.54</td>
<td>-0.11</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity*TurnDistance</td>
<td>-0.17</td>
<td>0.19</td>
<td>-0.53</td>
<td>0.22</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity*Politeness</td>
<td>-0.00</td>
<td>0.15</td>
<td>-0.30</td>
<td>0.30</td>
<td>1.00</td>
</tr>
<tr>
<td>TurnDistance*Politeness</td>
<td>-0.23</td>
<td>0.17</td>
<td>-0.59</td>
<td>0.11</td>
<td>1.00</td>
</tr>
<tr>
<td>Similarity<em>TurnDistance</em>Politeness</td>
<td>0.23</td>
<td>0.13</td>
<td>-0.02</td>
<td>0.50</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 7.13: Logistic regression output for experiment II [7] with Block order as a predictor.
7.7.3 Experiments I and II [6–7]: Sample experimental items

(a) First three turns.

(b) Last turn.

Figure 7.19: Talk item for experiments I and II [6–7] (1 intervening turn, perfect similarity, neutral politeness).
Discussion about the time of next week’s meeting

I have saved Wednesday 2pm in my phone.

Discussion about the time of next week’s meeting

I wrote down Tuesday 3pm.

Discussion about the time of next week’s meeting

I don’t know.

Discussion about the time of next week’s meeting

My calendar says Monday 4pm.

(a) First three turns.

(b) Last turn.

Figure 7.20: Talk item for experiments I and II [6–7] (3 intervening turns, perfect similarity, neutral politeness).
(a) Mismatch example for zero intervening turns and the neutral politeness condition.

(b) Mismatch example for three intervening turns and the impolite condition.
List of tables and figures
## List of Tables

3.1 Experiment 1 [2]: Raw utterance counts ........................................... 43  
3.2 Experiment 1 [2]: Evidentiality by utterance ..................................... 44  
3.3 Experiment 1 [2]: Categorical model output ...................................... 46  
3.4 Experiment 2 [3]: Evidentiality by utterance ..................................... 49  
3.5 Experiment 2 [3]: Binary logistic model output ................................ 50  
4.1 Production experiment [4]: Raw utterance counts ................................. 79  
4.2 Production experiment [4]: Evidentiality ratings by utterance and scenario 2 79  
4.3 Production experiment [4]: Categorical model output ......................... 81  
4.4 Production experiment [4]: Categorical model output with interaction .... 83  
4.5 Production experiment [4]: LOO approximation ................................... 85  
4.6 Comprehension experiment [5]: Sum coding for utterance .................... 90  
4.7 Comprehension experiment [5]: Speaker confidence ratings for each utterance 91  
4.8 Comprehension experiment [5]: Beta model output ............................ 92  
4.9 Comprehension experiment [5]: Beta model output with interaction ....... 94  
4.10 Comprehension experiment [5]: LOO approximation .......................... 95  
5.1 Experiment I [6]: Raw additive counts ........................................... 123  
5.2 Experiment I [6]: Logistic regression output ....................................... 125  
5.3 Experiment I [6]: LOO approximation ............................................. 127  
5.4 Experiment II [7]: Logistic regression output ................................... 134  
5.5 Experiment II [7]: LOO approximation ............................................. 135  
5.6 Experiment II [7]: Contrast coding for three-level Similarity ............... 136  
5.7 Experiment II [7]: Logistic regression output II ................................ 136  
7.1 Experiment [1]: Experimental items .................................................. ix  
7.2 Experiment 1 [2]: Categorical model output .................................... xviii  
7.3 Experiment 2 [3]: Categorical model output .................................... xix
7.4 Production experiment [4]: Categorical model output ................................ xx
7.5 Production experiment [4]: Categorical model output with order effect .......................... xxi
7.6 Comprehension experiment [5]: Beta model output with uninformative priors ............... xxvii
7.7 Comprehension experiment [5]: Beta model output with interaction ............................ xxvii
7.8 Experiment I [6]: Logistic regression output with wider priors for intercept .................. xxxii
7.9 Experiments I and II [6–7]: Sum-coding for factor Block order ................................. xxxiii
7.10 Experiment I [6]: Logistic regression output with Block order as a predictor ................ xxxiii
7.11 Experiment II [7]: Logistic regression output with wider priors I ............................... xxxiv
7.12 Experiment II [7]: Logistic regression output with wider priors II ............................ xxxiv
7.13 Experiment II [7]: Logistic regression output with Block order as a predictor ............ xxxv
List of Figures

2.1 Experiment [1]: Posterior distribution and 95% CrI: Objection rate ............... 24
3.1 Experiment 1 [2]: Box plot of evidentiality ratings for each formulation ........ 45
3.2 Experiment 2 [3]: Box plot of evidentiality ratings for each formulation ........ 49
3.3 Experiment 2 [3]: Prediction plot .................................................. 51
3.4 Experiment 2 [3]: Evidentiality by Participant ................................. 53
4.1 Production experiment [4]: Briefing item (suspect Emily Brown, strong evidence) 72
4.2 Production experiment [4]: Interrogation item (suspect Emily Brown, weak evidence) 73
4.3 Production experiment [4]: Control item (Briefing) ............................ 75
4.4 Production experiment [4]: Box plot of evidentiality ratings for each formulation .. 80
4.5 Production experiment [4]: Prediction plot 1 .................................... 81
4.6 Production experiment [4]: Prediction plot 2 .................................... 82
4.7 Production experiment [4]: Effect Difference Plot ............................... 84
4.8 Comprehension experiment [5]: Box plot of speaker confidence ratings for each formulation .......................................................... 92
4.9 Comprehension experiment [5]: Prediction plot .................................. 93
4.10 Comprehension experiment [5]: Effect Difference Plot ......................... 95
5.1 Experiments I and II [6–7]: Context pictures ........................................ 114
5.2 Experiments I and II [6–7]: Order item (perfect similarity, neutral politeness) .... 116
5.3 Experiments I and II [6–7]: Last order turn (perfect similarity, neutral politeness) . 117
5.4 Experiments I and II [6–7]: Last order turn (perfect similarity, reduced politeness) . 118
5.5 Experiments I and II [6–7]: Attention check context ............................. 119
5.6 Experiments I and II [6–7]: Attention check ....................................... 120
5.7 Experiment I [6]: Relative frequency of choosing additives ....................... 124
5.8 Experiment I [6]: Prediction Plot ......................................................... 126
5.9 Experiment I [6]: Raw counts of additives and *same* ............................ 129
5.10 Experiment II [7]: Relative frequency of additive production .................................. 133
5.11 Experiment II [7]: Prediction plot .......................................................... 135

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Item Type</th>
<th>Suspect</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>Experiments [2–3]: Briefing item</td>
<td>Emily Brown</td>
<td>weak evidence</td>
</tr>
<tr>
<td>7.2</td>
<td>Experiments [2–3]: Briefing item</td>
<td>Emily Brown</td>
<td>strong evidence</td>
</tr>
<tr>
<td>7.3</td>
<td>Experiments [2–3]: Interrogation item</td>
<td>Emily Brown</td>
<td>weak evidence</td>
</tr>
<tr>
<td>7.4</td>
<td>Experiments [2–3]: Interrogation item</td>
<td>Emily Brown</td>
<td>strong evidence</td>
</tr>
<tr>
<td>7.5</td>
<td>Experiments [2–3]: Briefing item</td>
<td>Johanna Smith</td>
<td>weak evidence</td>
</tr>
<tr>
<td>7.6</td>
<td>Experiments [2–3]: Briefing item</td>
<td>Johanna Smith</td>
<td>strong evidence</td>
</tr>
<tr>
<td>7.7</td>
<td>Experiments [2–3]: Interrogation item</td>
<td>Johanna Smith</td>
<td>weak evidence</td>
</tr>
<tr>
<td>7.8</td>
<td>Experiments [2–3]: Interrogation item</td>
<td>Johanna Smith</td>
<td>strong evidence</td>
</tr>
<tr>
<td>7.9</td>
<td>Production experiment [4]: Prediction plot Inter Interaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.10</td>
<td>Production experiment [4]: Briefing item</td>
<td>Emily Brown</td>
<td>weak evidence</td>
</tr>
<tr>
<td>7.11</td>
<td>Production experiment [4]: Briefing item</td>
<td>Emily Brown</td>
<td>strong evidence</td>
</tr>
<tr>
<td>7.12</td>
<td>Production experiment [4]: Interrogation item</td>
<td>Emily Brown</td>
<td>weak evidence</td>
</tr>
<tr>
<td>7.13</td>
<td>Production experiment [4]: Interrogation item</td>
<td>Emily Brown</td>
<td>strong evidence</td>
</tr>
<tr>
<td>7.14</td>
<td>Comprehension experiment [5]: Briefing item</td>
<td>Emily Brown, ‘know’</td>
<td></td>
</tr>
<tr>
<td>7.15</td>
<td>Comprehension experiment [5]: Interrogation item</td>
<td>Emily Brown, ‘believe’</td>
<td></td>
</tr>
<tr>
<td>7.16</td>
<td>Comprehension experiment [5]: Interrogation item</td>
<td>Johanna Smith, ‘know’</td>
<td></td>
</tr>
<tr>
<td>7.17</td>
<td>Comprehension experiment [5]: Interrogation control item</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.18</td>
<td>Comprehension experiment [5]: Interrogation control item</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.19</td>
<td>Experiments I and II [6–7]: Talk item</td>
<td>Emily Brown, 1 intervening turn, perfect similarity, neutral politeness</td>
<td></td>
</tr>
<tr>
<td>7.20</td>
<td>Experiments I and II [6–7]: Talk item</td>
<td>Emily Brown, 3 intervening turns, perfect similarity, neutral politeness</td>
<td></td>
</tr>
</tbody>
</table>