Believe is Strong but Subjective: Experimental Evidence from Hedging[1]

Todor KOEV — *University of Konstanz*
Cory BILL — *University of Konstanz*
Maryam MOHAMMADI — *University of Konstanz*

Abstract. This paper contributes to the debate concerning whether the attitude verb *believe* has a weak or a strong semantics. According to Hawthorne et al. (2016), *believe* is weak and akin to the probability operator *likely*, typically receiving an "agent finds it more likely than not" interpretation. Alternatively, Koev (2019) proposes that *believe* conveys high certainty but qualifies this certainty as subjective or lacking evidence, in contrast with modals that convey high objective certainty, like *sure* (see also Lyons 1977; Kratzer 1981; Nuyts 2001; Papafragou 2006; Portner 2009). We focus on the use of *believe* as a hedge (e.g. *I believe the Giants will win, but I'm not sure they will*) as allegedly the most convincing argument for the weak view, and argue that, in fact, it favors the strong-but-subjective view. We show experimentally that the availability of the hedging use of *believe* is affected by certain grammatical and discourse factors. Experiment 1 reveals that participants rate hedging sentences with combinations of third person/past tense/embedded features as less natural than canonical first person/present tense/main clause forms. In turn, Experiment 2 reveals that hedging sentences with at-issue prejacents are judged as more natural than sentences in which the belief component is at-issue. The observed variability posits a challenge to the weak view, which establishes a purely logical contrast in modal strength between likelihood vs. certainty. However, it is in line with the strong-but-subjective view, which establishes a contrast in modal content between certainty without evidence vs. certainty with evidence and predicts a more restricted distribution of the hedging reading.

Keywords: belief, modality, strength, gradability, subjectivity, experimentation.

1. The Hintikkan orthodoxy and the issue of strength

Ever since Hintikka (1969), the standard view in formal semantics has been that the attitude verb *believe* expresses universal quantification over possibilities (e.g., world-time pairs). More specifically, this view states that the **prejacent** (the complement of *believe*) is true in all of the agent's doxastic alternatives. This is usually formalized as in (1), where $Dox_{x,i}$ stands for the set of x's doxastic alternatives at i, i.e., the set of indices compatible with everything x believes in i.

(1) $[\![\text{believe}]\!]^i \;=\; \lambda p \lambda x \,.\, \forall i' \in Dox_{x,i} : p(i')$

One outstanding issue with the standard view is that it makes no predictions about how strongly – i.e., fully or partially – *believe* commits the agent to the prejacent. Although the prejacent is stated to be true across all doxastic alternatives, the universal force alone does not entail a specific degree of certainty, the reason being that it remains unclear how the agent is linked to the domain of quantification itself. Since *Dox* is defined as the set of indices compatible with

[1] We would like to thank Elsi Kaiser and the audiences at *Sinn und Bedeutung* 25 for stimulating feedback. We would also like to thank each other. The research on his paper was funded by DFG grant KO 5704/1-1.

everything the agent "believes" (at the relevant index), the issue of strength is merely pushed into the metalanguage.

2. Two views on the strength of *believe*

There are two views on the strength of *believe* proposed in the literature. One view is that *believe* carries a weak (or non-strong) modal force that is comparable to the force of *likely* (Hawthorne et al. 2016). Another view is that *believe* carries a strong but subjective force, unlike *sure*, which is strong and objective (Koev 2019). Before spelling out these competing analyses and in order to facilitate the comparison, we introduce a gradable semantics for *believe*, which is presupposed by both views. We also introduce a distinction between subjective and objective epistemic modality, which lies at the heart of the latter, strong-but-subjective view.

2.1. *Believe* as a gradable predicate

There is robust empirical evidence that *believe* is a gradable predicate. It can participate in a range of constructions that make reference to degrees of belief, e.g. comparatives (2a) and equatives (2b), if mediated through gradable adverbs like *strongly*. It can also be modified by minimality (3a), maximality (3b), and proportional modifiers (3c). The examples below naturally occur on the web.

(2) a. He believes more strongly than I do that the organization of the executive branch of the federal government matters a great deal.
b. Each [farmer] believes as strongly as the other that his crops will not survive another week without water, and each cares as much as the other about the survival of his crops.

(3) a. Atticus partially believes that prejudice exists because people do not understand each other, he wants to teach his children that they need to consider ideas from everyone's point of view.
b. I strongly believe that life is too short to eat mediocre meals.
c. This has taken me lots of research to come to this conclusion, but I believe 95 percent that it is.

The gradability of *believe* can be captured by a semantics similar to that proposed for gradable adjectives (Cresswell 1976; von Stechow 1984; Kennedy 1999; a.m.o.). We can assume that *believe* encodes a measure function μ (parameterized by agents, worlds, and times) and makes available a degree argument, in addition to its propositional and individual arguments. The entry below states that the belief agent's certainty in the prejacent meets some norm.

(4) $[\![\text{believe}]\!]^{w,t} = \lambda p \lambda d \lambda x \,.\, \mu_{x,w,t}(p) \geq d$

The degree argument of *believe* is always filled by degree morphology. In "positive" forms, which lack overt such morphology, the norm or standard of comparison is supplied by a covert POS morpheme (Cresswell 1976). Following Kennedy and McNally (2005), we assume that POS is sensitive to the modified predicate and may also be sensitive to the discourse context. It provides different values, depending on whether the gradable predicate it combines with is **absolute** (minimum-degree like *wet* or maximum-degree like *full*) or **relative** (like *tall*). In cases where *believe* is overtly modified, e.g. by the (minimum) degree modifier *partially*, the standard of comparison is a specific value on the relevant scale. In the case of *partially*, it is the minimal degree of belief.

The following examples in (5) and (6) give a sense of how POS and *partially* are composed with a VP headed by *believe*. The parameter C is a contextually supplied comparison class of appropriate objects, $\mathbf{standard}(d,P,C)$ determines the degree d relative to the predicate P and the comparison class C, and $\mathbf{min}(s_P)$ refers to the minimal value of the P-scale s.[2]

(5) $[_{\text{DegP}}$ POS $[_{\text{VP}}$ believe $[_{\text{TP}}$ it is raining$]]]$

a. $[\![[_{\text{VP}} \text{believe } [_{\text{TP}} \text{it is raining}]]]\!]^{C,w,t} = \lambda d' \lambda y \,.\, \mu_{y,w,t}([\![\text{it is raining}]\!]^C) \geq d'$

b. $[\![\text{POS}]\!]^{C,w,t} = \lambda P \lambda x \,.\, \exists d[\mathbf{standard}(d,P,C) \wedge P(d)(x)]$

c. $[\![[_{\text{DegP}} \text{POS } [_{\text{VP}} \text{believe } [_{\text{TP}} \text{it is raining}]]]]\!]^{C,w,t}$

$$= \lambda x \,.\, \exists d \begin{bmatrix} \mathbf{standard}(d, \lambda d' \lambda y \,.\, \mu_{y,w,t}([\![\text{it is raining}]\!]^C) \geq d', C) \wedge \\ \mu_{x,w,t}([\![\text{it is raining}]\!]^C) \geq d \end{bmatrix}$$

(6) $[_{\text{DegP}}$ partially $[_{\text{VP}}$ believe $[_{\text{TP}}$ eating pizza is healthy$]]]$

a. $[\![[_{\text{VP}} \text{believe } [_{\text{TP}} \text{eating pizza is healthy}]]]\!]^{C,w,t}$
$= \lambda d' \lambda y \,.\, \mu_{y,w,t}([\![\text{eating pizza is healthy}]\!]^C) \geq d'$

b. $[\![\text{partially}]\!]^{C,w,t} = \lambda P \lambda x \,.\, \exists d[d > \mathbf{min}(s_P) \wedge P(d)(x)]$

c. $[\![[_{\text{DegP}} \text{partially } [_{\text{VP}} \text{believe } [_{\text{TP}} \text{eating pizza is healthy}]]]]\!]^{C,w,t}$
$= \lambda x \,.\, \exists d[d > \mathbf{min}(s_P) \wedge \mu_{x,w,t}([\![\text{eating pizza is healthy}]\!]^C) \geq d]$

(5) amounts to the set of individuals whose certainty in it being raining meets or exceeds the norm determined by the **standard** function. (6) amounts to the set of individuals whose certainty in the proposition that eating pizza is healthy is above the minimal degree of belief.[3]

2.2. Subjective vs. objective epistemic modality

Lyons (1977) points out that epistemic modality comes in two flavors: **subjective** vs. **objective**. This is illustrated with epistemic *must*, which, as shown in (7), can take on either reading.[4]

(7) Alfred must be unmarried. (Lyons 1977: 791–792)

a. Subjective: I (confidently) infer that Alfred is unmarried.

b. Objective: In the light of what is known, it is necessarily the case that Alfred is unmarried.

According to Lyons, subjective modality expresses an opinion while objective modality is based on evidence. The former essentially serves as an illocutionary modifier: it manipulates the degree of public commitment to the prejacent. The function of the latter is instead truth-conditional. It contributes to the propositional content of the sentence.

Kratzer (1981) attempts to model this distinction by assuming that the two readings share the same domain of quantification (or "modal base") but differ in how this domain is structured

[2] We do not take a stand on how the standard value differs across worlds or times. So there are no world/time parameters on the **standard** function.

[3] See Koev (2019) for more formal details.

[4] We use the term "reading" informally here. We do not claim that epistemic modals are lexically ambiguous between a subjective and an objective interpretation. See the Conclusion for brief discussion.

(i.e., they differ in what the “ordering source” is). She offers the following illustration (the original example is in German, see Kratzer 1981: 307).

(8) *Lenz, who often has bad luck, is going to leave the Old World by boat today, on Friday 13. On hearing about this, someone utters*:
  a. Probably the boat will sink.
  b. It is probable that the boat will sink.

According to Kratzer, the claim in (8a) is subjective. It is based on superstitions and cannot be defended on objective grounds, hence it requires a subjective (stereotypical) background as an ordering source. In contrast, the claim in (8b) is objective. It is based on established facts about the boat, the technical equipment or the weather, and can be defended on objective grounds. It requires an objective (stereotypical) background as an ordering source. Intuitively, this contrast has something to do with the fact that (8a) contains the modal adverb *probably* while (8b) contains the modal adjective *probable*.

Nuyts (2001) and Papafragou (2006) put a different spin on the above distinction. For them, the distinction is not about the quality of the evidence but about its status, i.e. whether it is shared. Subjective modality involves evidence only known to the speaker while objective (or “intersubjective”) modality involves evidence shared among all speech participants. Portner (2009: 4.2) demonstrates that both Lyons’ and Nuyts/Papafragou’s characterizations can be modeled within a Kratzerian modal semantics.

We follow Lyons’ original characterization of the contrast in subjectivity as being about the quality of the evidence. We would like to also stress two points about our understanding of this contrast. The first point is that the subjective/objective distinction is about modal content and is orthogonal to the implied degree of strength, which is about modal force. Most importantly, a subjective use does not entail a low degree of certainty in the prejacent. The second point is that modal items may lexicalize a particular flavor of epistemic modality. Thus, mental state predicates like *believe*, *think* or *doubt* are inherently subjective, so they invariably convey private opinions and may serve as hedges on public commitments. Nuyts (2001: 390–391) touches on both of these points as he writes:

> The mental state predicates systematically express subjectivity. [...] Because the mental state predicates are inherently subjective, they are frequently used as mitigating or hedging devices [...]. In such uses, it is usually quite obvious that speakers are absolutely certain about or convinced of what they are saying, but by using the mental state predicate they suggest that they are voicing a tentative and personal opinion which may be wrong, thus ‘officially’ leaving room for another opinion or for a reaction on the part of the hearer.

2.3. Weak Believers

One view on *believe* is that the certainty it invokes exceeds some contextually determined threshold. This threshold is typically 50% but it can shift somewhat when there are several alternatives to the prejacent. This view is espoused in Hawthorne et al. (2016) and we call it **Weak Believers** (WB). Making use of the gradable semantics for *believe* proposed in (4), this view can be formalized as follows.[5]

(9) Weak Believers
$(\llbracket \text{POS} \rrbracket^{C,w,t}(\llbracket \text{believe} \rrbracket^{C,w,t}(p)))(x)$ iff $\mu_{x,w,t}(p) > \theta_{bel}$, where typically $\theta_{bel} = 0.5$

Roughly, WB states that *believe* parallels *likely* in that the threshold value is taken from the middle of the relevant scale (Yalcin 2010; Lassiter 2017). Outside the modal domain, *believe* bears similarities to proportional quantifiers like *most* or *more than half* (Barwise and Cooper 1981; Hackl 2009; Solt 2016) in that it has some sort of a "majority" interpretation.

2.4. Strong Subjective Believers

Koev (2019) argues for a different view, according to which *believe* conveys maximal certainty but this certainty is subjective, in the sense of lacking (reliable) evidence. In this, *believe* differs from epistemic modals like *sure*, which we suggest encodes high but objective certainty that is backed by evidence. The core intuition this characterization tries to capture is that subjective modality weakens the public commitments of the agent without necessarily lowering her certainty level. In order to distinguish between these two kinds of epistemic modality, we split the generic probability function μ from above into two separate functions: *Cr* measures "credences" or subjective certainty while *Pr* measures objective certainty.[6] We dub this view **Strong Subjective Believers** (SSB) and spell it out as follows.

(10) Strong Subjective Believers
$(\llbracket \text{POS} \rrbracket^{C,w,t}(\llbracket \text{believe} \rrbracket^{C,w,t}(p)))(x)$ iff $Cr_{x,w,t}(p) = 1$

SSB states that the strength of *believe* is on par with *certain*, as both elements pick their threshold value from the top of the scale. In the non-modal domain, *believe* is comparable to universal quantifiers in that it receives a "maximality" interpretation.

3. Empirical evidence for WB and SSB

Several arguments have been put forward in support of WB or SSB. This section lists three arguments for each view. The third argument for WB, i.e. the possible use of *believe* as a hedge, is further investigated through a set of two experiments presented later in the paper.

3.1. Arguments for WB

One argument for WB comes from neg-raising, a phenomenon whereby a matrix negation is interpreted as though it takes scope inside the embedded clause (Bartsch 1973; Horn 1989; Gajewski 2007; Romoli 2013; Homer 2015). *Believe* is a classic neg-raising predicate, so that *x*

[5] Hawthorne et al. add a second condition which requires that the prejacent be significantly more likely than any of its alternatives: $\mu_{x,w,t}(q) \gg \mu_{x,w,t}(p)$, for all alternatives q of p. In the case of a binary choice, this condition boils down to $\mu_{x,w,t}(p) \gg \mu_{x,w,t}(\neg p)$, which entails that the likelihood of the prejacent is significantly greater than 0.5. This second condition does not substantially change the account and it is ignored here.

[6] Notice that the two measures and corresponding certainty types are linked. Assuming that the speaker is sincere, one can think of *Pr* as a more conservative version of *Cr*. That is, if a speaker is publicly committed to some proposition to a given degree, her subjective certainty in that proposition will meet that degree: $Pr_{x,w,t}(p) \leq Cr_{x,w,t}(p)$, for all agents x and propositions p. It is easy to think of *Cr* as encoding degrees of belief and of *Pr* as encoding degrees of knowledge, but we will not push this claim too hard.

doesn't believe p often comes to mean *x believes not p*. Since *believe* licenses neg-raising, one might wonder whether this property is characteristic of modals of a particular strength. Indeed, Hawthorne et al. (2016) hypothesize that neg-raising occurs with weak predicates (like *think* or *want*) but not with strong predicates (like *know* or *need*). Since *believe* shares this property with the former group, they conclude that it must have a weak semantics.

One issue with this argument is that the alleged link between modal strength and neg-raising is not very tight. Horn (1989: ch.5) draws a distinction between (properly) weak scalars like *possible* or *allowed*, mid-scalars like *likely*, and strong scalars like *certain* or *necessary*, with an eye on their neg-raising behavior.[7] Specifically, he adduces cross-linguistic evidence to argue for the following generalization: weak scalars never license neg-raising, mid-scalars typically do, and strong scalars may or may not license it. On this taxonomy, according to WB, *believe* is a mid-scalar, so it is unsurprising that it licenses neg-raising. However, if WB is wrong and *believe* is in fact a strong scalar, then – according to Horn – it is still possible that it would license neg-raising. Therefore, the fact that *believe* licenses neg-raising can only stand as an argument in favor of it being a mid- or strong scalar and against it being a weak scalar. And by extension, this argument cannot be used to adjudicate between WB and SSB.

A second argument for WB comes from what one may call "gradation sequences". If *believe* is weak, it should be possible to strengthen it by using a modal with a stronger force, such as *know*. This is indeed borne out.

(11) Scientists believe there is water on Mars. In fact, they know it.

While this is a plausible analysis of (11), we note that assuming a stronger modal force for *know* than that for *believe* is not the only way in which such data can be captured (cf. Percus 2006; Chemla 2008; Sauerland 2008; Schlenker 2012). That is, it is relatively uncontroversial that *know* is "stronger" than *believe* in at least two other respects: (i) *know* is a factive verb that entails its prejacent, and (ii) *know* entails that the attitude holder has appropriate evidence for the prejacent. So the final clause in (11) may actually involve a raising of the commitment of the speaker towards the prejacent without ascribing a raised level of certainty to the attitude holder (the group of scientists in question).

The third and seemingly strongest argument in favor of WB comes from the use of *believe* as a hedge. An example of that use is presented in (12). One simple explanation of such uses is that *believe* expresses high but non-maximal certainty, hence the sentence is not contradictory. This is fully in line with WB.

(12) I believe it's raining, but I'm not sure it's raining. (Hawthorne et al. 2016: 1395)

As we will outline in Section 4, SSB is also able to account for cases like (12). But it involves additional assumptions that WB does not need to make.

[7] Notice that Horn's "weak scalars" and "mid-scalars" are collapsed by Hawthorne et al. into the one category of "weak" modals.

3.2. Arguments for SSB

One basic argument for SSB is that belief attributions systematically lack quantity implicatures to uncertainty. That is, (13a) does not implicate (13b).[8]

(13) a. Kamala believes that America needs universal health care.
b. $\not\rightsquigarrow$ Kamala is not certain that America needs universal health care.

The systematic lack of such implicatures would be surprising if *believe* had a weak semantics. But it is as expected if *believe* conveys full (subjective) certainty.

A second argument for SSB comes from the way *believe* interacts with conjunction. A modal operator M is said to be **closed under conjunction** when the following holds: $M(p) \wedge M(q)$ entails $M(p \wedge q)$. Now, notice that strong modals obey this property while non-strong modals do not.[9] That is, a sentence like (14a) entails (14b), therefore the strong modal *certain* is closed under conjunction. In contrast, given the context in (15), the sentences in (15a) and (15b) do not entail (15c), therefore the non-strong modal *probably* is not closed under conjunction.

(14) a. It's certain that Sean is in Rome and it's certain that he is catholic.
b. $\models$ It's certain that Sean is in Rome and that he is catholic.

(15) *Each week Jack spends (in no particular order) 3 nights at the local pub and gets drunk, 2 nights at the same pub but stays sober, and 2 nights at home where he also gets drunk. On a given night, I say*:
a. Jack is probably at the pub. True (chance = 5/7)
b. Jack is probably drunk. True (chance = 5/7)
c. Jack is probably at the pub drunk. False (chance = 3/7)

As for *believe*, it lines up with *certain* in that it is also closed under conjunction relative to its prejacent argument, as (16) demonstrates. Given that only strong modals share this property, we take this as evidence that *believe* is strong.

(16) a. Ron believes Mia is hawt and he also believes she is going to marry him.
b. $\models$ Ron believes that Mia is hawt and that she is going to marry him.

Example (16) merely establishes the empirical point that *believe* pairs up with *certain* and differs from *probably* in how it interacts with conjunction. Notice, in addition, that the conjunction closure of *believe* is a direct consequence of SSB but it is not warranted by WB.[10]

The third argument for SSB that we provide here is a bit more involved and has to do with the gradability properties of *believe*. When occurring outside a degree construction, gradable predicates of different kinds pick different standards of comparison. Unger (1971) thus distinguished between relative adjectives, whose standard is vague and is taken from the middle of the scale, and absolute adjectives, whose standard is fixed as the minimum or the maximum of

[8] All data in this section are taken from Koev (2019).

[9] The reader can verify that very weak modals like *possible* are not closed under conjunction either. We skip the relevant data for reasons of space.

[10] This is so for the following reason. If $\mu(p) = 1$ and $\mu(q) = 1$, then necessarily $\mu(p \wedge q) = 1$, for all probability measures μ. However, let $\theta = 0.5$ and let p and q be probabilistically independent. Then, if $\mu(p) = 0.6$ and $\mu(q) = 0.7$, both p and q meet μ. But $\mu(p \wedge q) = \mu(p) \times \mu(q) = 0.42$, so $p \wedge q$ does not meet μ. In general, any threshold value lower than 1 destroys the conjunction closure property.

the scale. Kennedy and McNally (2005) convincingly argue that the relative/absolute distinction boils down to differences in scale structure. Relative adjectives are associated with an open scale that lacks endpoints, so they generally cannot be modified by minimality, maximality, or proportional adverbs (cf. **slightly tall*, **completely tall*, **half tall*). In contrast, absolute adjectives are associated with lower-closed, upper-closed, or totally closed scales, and thus accept minimality, maximality, or proportional modifiers, respectively (cf. *slightly empty*, *completely empty*, *half empty*). Now, the main thing to notice is that sentences with *believe* accept minimality, maximality, and proportional modifiers, as already established in (3). This suggests that *believe* is associated with a totally closed scale, i.e. one that has a minimum and a maximum. It must then pick an absolute (i.e., 0 or 1), not a relative standard of comparison (e.g., 0.5), just like SSB predicts.

4. Divergent predictions about the hedging data

The previous section has shown that, in general, SSB is better supported by the empirical evidence than WB. However, the hedging data stands out as one very simple and intuitive argument in favor of WB. So let us zero in on the hedging use of *believe* and see how the two rivaling views measure up to it.

Consider (17) below. As discussed in Section 3.1, WB accounts for such data in terms of modal strength: the speaker assigns non-maximal certainty to the prejacent. SSB offers a very different explanation: it suggests that while the speaker expresses a maximal subjective certainty regarding the prejacent, they are also unwilling to publicly commit to it due to lack of (sufficient) evidence, thus conveying lower objective certainty. The two views are juxtaposed in (17a) vs. (17b) below. We assume that *sure* encodes objective probabilities and use the following abbreviations in the schematic representations: $\triangle =$ it is likely, $\Box =$ it is certain, $s =$ subjective, $o =$ objective.

(17) I believe the Giants will win, but I'm not sure they will.
a. WB: $\mu_{x,w,t}([\![\text{the Giants will win}]\!]^C) > \theta_{bel} \wedge \mu_{x,w,t}([\![\text{the Giants will win}]\!]^C) < 1$
$\triangle p \wedge \neg \Box p$
b. SSB: $Cr_{x,w,t}([\![\text{the Giants will win}]\!]^C) = 1 \wedge Pr_{x,w,t}([\![\text{the Giants will win}]\!]^C) < 1$
$\Box_s p \wedge \neg \Box_o p$

Although each view takes the hedging data in its stride, WB clearly wins out on simplicity, because it does not have to distinguish between two different flavors of epistemic modality and corresponding probability measures. However, this is not all. These explanations diverge in at least two further predictions that we investigate over the course of two experiments.

One point of divergence has to do with the grammatical form of hedging sentences. WB draws a purely logical contrast, rooted in a single probabilistic measure and comparable to "likely but not certain". WB predicts, therefore, that such sequences are licit across the board. Specifically, it should make no difference whether (17) occurs in the first or third person, in the present or past tense, unembedded or embedded under hypotheticals like *suppose*. In turn, SSB draws a contrast between stated private beliefs and incurred discourse commitments, and this requires a salient speech context. Thus, first person/present tense/unembedded variants of (17) – as most closely tied to the utterance context – are expected to be more natural than third person/past tense/embedded forms, which are further detached from the utterance context and the subjective vs. objective contrast is harder to draw.[11]

The second domain in which these accounts make divergent predictions is that of discourse structure. According to WB, (17) should be fine regardless of whether what is at-issue is the prejacent (i.e., answering QUD$_1$ = *Will the Giants win?*) or the belief component itself (e.g., answering QUD$_2$ = *Why did you bet on the Giants?*). That is, if the prejacent is at-issue, since both clauses qualify its likelihood, they will contribute to the same QUD. Therefore, the combination of these clauses under QUD$_1$ is expected to proceed smoothly. Now, if what is at-issue is instead the belief component, as under QUD$_2$, then again both clauses will be perceived as addressing the same question, although here the focus will be on the agent's attitude toward the prejacent. Since (17a) invokes a single measure of certainty and the two clauses are of the same shape, they are not expected to differ in discourse status under either construal.[12]

Moving on to the SSB analysis in (17b), the sentence is predicted to be natural when answering QUD$_1$, because when the prejacent is at-issue the discourse topic invokes objective probabilities, so the speaker has to hedge in order to avoid too strong a commitment. In contrast, (17b) is expected to be less natural with QUD$_2$. In this latter case, QUD$_2$ is asking about a personal estimate (or subjective certainty) while the speaker additionally invokes objective certainty in the follow-up clause, a move that is likely to feel irrelevant in this context. Notice that we do not expect a sharp contrast here, because new QUDs can easily be accommodated in most circumstances. But we do expect some effect in the predicted direction. That is, while in both cases the speaker pivots to a QUD which involves a different flavor of certainty, pivoting from objective to subjective certainty (as occurs under QUD$_1$) is expected to be more natural than pivoting in the opposite direction (as occurs under QUD$_2$).

The common thread that runs through both sets of divergent predictions is that SSB implies a more restricted distribution of the hedging use of *believe* than does WB. We tested these predictions in two experiments and found that the naturalness of hedging sentences is sensitive to the noted grammatical and discourse factors, as expected by SSB but not by WB. We turn to these experiments now.

5. Experiment 1: Grammatical factors

Experiment 1 tested the noted divergent predictions regarding the effects of grammatical factors on the acceptability of sentences in which *believe* is used as a hedge. The experiment had a 2×2×2 structure, with the factors Person (first vs. third), Tense (present vs. past), and Position (main clause vs. embedded under *suppose*). We also collected comments from our participants to check their intuitions about sentence acceptability and to help identify any issues with certain

[11] We focus on person, tense, and position as three grammatical features that may lead to variable judgements for two main reasons. First, the proposed semantics for *believe* in (4) already makes these parameters available, so we expect them to have an impact on acceptability, at least under SSB. Moreover, there is an empirical parallel for this kind of variation with other types of epistemic contradictions. It is known, for example, that the acceptability of Moore-paradoxical sentences are sensitive to precisely these three features, although the effect is in the opposite direction (Moore 1993: ch.12). That is, while *It's raining but I don't believe it* is odd, each of *It's raining but Mia doesn't believe it*, *It was raining but I didn't believe it*, and *Suppose it's raining but I don't believe it* are fine.

[12] It is still possible to claim that the contribution of epistemic modals is generally more difficult to be construed as being at-issue than the prejacent (cf. Papafragou 2006; but see Simons 2007). However, as discussed in the following paragraph, SSB independently predicts this contrast and thus has more predictive power.

items or the experiment overall.

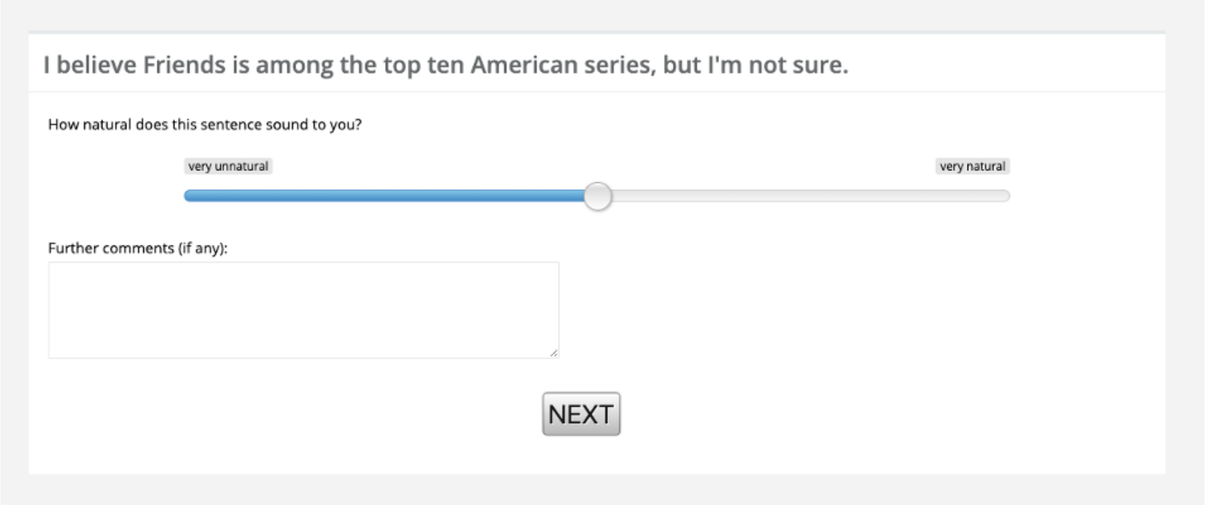
5.1. Participants

We recruited 96 participants from Amazon Mechanical Turk. The participants were required to be residing in the US (i.e. be accessing the experiment for a US IP address) and to be over 18 years of age. Informed consent was obtained from all participants and they were compensated with a small payment. Seven participants were excluded based on the exclusion criterion outlined below, reducing the number of participants to 89.

5.2. Procedure

We used a sentence acceptability task, in which participants rated sentences for naturalness on a continuous scale from very unnatural (0) to very natural (100). The experiment started by presenting participants with a short description of the task, which included an explicit direction not to judge the sentences based on real-world plausibility but on naturalness. The participants were not told the purpose of the study.

The experiment was presented through `testable.org`, a platform designed for running online experiments. As shown in Figure 1, for each item participants were presented with a target sentence, a slider to rate its acceptability, and a textbox which they could (optionally) use to share any further thoughts about the sentence.



I believe Friends is among the top ten American series, but I'm not sure.

How natural does this sentence sound to you?

Further comments (if any):

Figure 1: Sample trial in Experiment 1.

5.3. Materials

We constructed 16 core sentences and crossed them with our three factors (i.e. Person, Tense, and Position), resulting in eight distinct conditions and a total of 128 test sentences. All targets were of the form $believe_x\, p \wedge \neg sure_x\, p$, with the second conjunct reinforcing the weak/subjective nature of *believe* by contrasting it with *sure*, which conveys strong (objective) certainty. For example, (18) presents one of the test sentences from the first person/present tense/main clause condition.

(18) I believe it's raining, but I'm not sure. (sample target)

In addition to the test sentences, the experiment contained 18 filler sentences, which were included to distract the participants from the predicate *believe* and to check that they were attending to the task. The filler sentences were of the same form as the test sentences but included different predicates, e.g. *guess*, *know*, *imagine*. Fillers were divided into "good" and "bad" based on their expected acceptability. While all the filler sentences were grammatically well-formed, bad fillers presented semantic contradictions whereas good fillers were meaningful and non-contradictory, as illustrated below.

(19) a. I guess she's from Canada, but I'm not sure. (good filler)
b. I know what you are saying is true, but it's false. (bad filler)

Participants who failed three or more of the bad fillers by rating them higher than 40% were excluded form the data-set based on the assumption that such responses reflected a lack of attention.

5.4. Design

Eight stimulus lists were constructed using a Latin square design, so that each list contained each of the 16 test sentences in only one of the eight conditions, thus a given list contained two sentences per condition. The items in each list were presented in a pseudo-randomized order and eighteen filler items were added in between test items. Each list started with two filler items, a good one followed by a bad one, as a warm-up. Participants were randomly assigned to one of the eight experimental lists. Each participant was presented with all 18 filler trials and all 16 experimental trials and each sentence was presented only once. The experiment duration was on average eight minutes.

5.5. Data analysis

The core results are summarized in Figure 2. We conducted the statistical analysis with R (R Core Team 2020), using the *lmer* function from the *lme4* package. We fitted a mixed effects linear regression model to the data, with the fixed effects of Position (Main vs. Emb), Tense (Pres vs. Past), and Person (1P vs. 3P), and all their interactions. Following Barr et al. (2013), we identified the optimal random effects structure via forward model selection guided by the **best path algorithm**.[13] This resulted in a model which included random intercepts for participant and item, as well as random by-participant slopes for Position.

We compared the full model to models without the different fixed effects and their various interactions. These model comparisons revealed significant main effects of Position ($\chi^2(1) = 69$, $p < 0.001$), Tense ($\chi^2(1) = 55, p < 0.001$), and Person ($\chi^2(1) = 8, p < 0.01$). We also found significant effects for the interactions between Position and Tense ($\chi^2(1) = 66, p < 0.001$), Position and Person ($\chi^2(1) = 13, p < 0.001$), and Position, Tense, and Person ($\chi^2(1) = 16$, $p < 0.001$). No significant effect was found for the interaction of Tense and Person ($\chi^2(1) = 0.02, p = 0.88$).

[13] This algorithm involves starting with a simple model (i.e. fixed-effects and random effect intercepts) and then, at each step, testing (via model comparison-based significance testing) for the potential inclusion of all random effects not currently in the model.

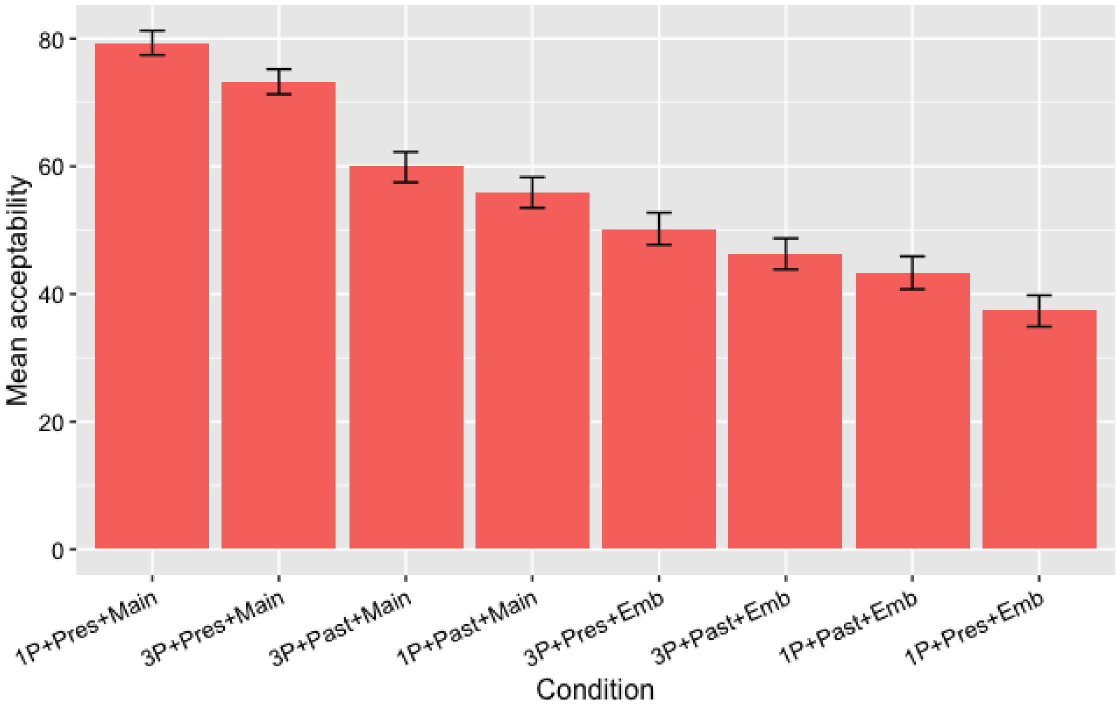


Figure 2: Average proportion of sentence acceptability ratings per condition in Experiment 1. Error bars indicate standard errors.

These significant effects reveal that all three factors affected participants' acceptability judgments of the test sentences to some extent. In the case of Tense and Position we find that, as predicted by SSB, participants rated as more acceptable present tense/main clause forms than past tense/embedded clause forms. However, in the case of Person, while we found a significant main effect, the direction of this difference was the opposite of what SSB predicts: third person forms were rated as overall more acceptable than first person forms. We discuss some possible explanations for this result and its implications below.

6. Experiment 2: Discourse factors

Experiment 2 was designed to test the divergent predictions of WB and SSB with regard to the acceptability of hedging sentences in different discourse contexts. Specifically, we compared the acceptability of such sentences in contexts where the prejacent was at-issue to those where the belief component itself was at-issue.

6.1. Participants

We recruited 62 participants from Amazon Mechanical Turk. As with Experiment 1, the participants were required to be residing in the US and to be over 18 years of age. Informed consent was obtained from all participants and they were compensated with a small payment. Fourteen participants were excluded based on the exclusion criterion outlined below, reducing the final number of participants to 48.

6.2. Procedure

The procedure was identical to Experiment 1 (see Section 5.2).

6.3. Materials

The experiment contained a single factor At-issueness with two levels: Prejacent vs. Belief. We constructed four core sentences, crossed them with this factor to create two experimental conditions and a total of eight test sentences. All test items involved question/answer pairs, where the questions determined the condition and the answers were of the form $believe_x\, p \wedge \neg\, sure_x\, p$. For example, (20) was an item in the Prejacent condition and (21) was an item in the Belief condition.

(20) A: Are Nike Zoom the best running shoes? (Prejacent condition)
B: I believe they are, but I'm not sure.

(21) A: Why did you buy Nike Zoom running shoes? (Belief condition)
B: I believe they are the best, but I'm not sure.

We also created six filler sentences, included to distract participants from the task at hand and to check that they were paying attention. Filler items contained polar or constituent questions and answers varied across attitude verb. As in Experiment 1, fillers were divided into "good" and "bad" ones, where the latter presented semantic contradictions. Participants who failed all three bad fillers by rating them higher than 50% were excluded from the final data-set.

6.4. Design

Two stimulus lists were constructed using a Latin square design, so that both lists contained each of the four core sentences in only one condition. Thus, a given list contained two sentences per condition. Both lists started with two filler items, a good one followed by a bad one, and then prejacent and belief conditions followed upon each other interspersed with filler items. Participants were randomly assigned to one of the two experimental lists. In total, each participant was presented with six filler items and four experimental items. The experiment duration was approximately four minutes.

6.5. Data analysis

We conducted the statistical analysis with R, using the *lmer* function from the *lme4* package. The theoretical method to identify the optimal model was the same as Experiment 1 (see Section 5.5). The optimal random effect structure was one which included both by-participant and by-item intercepts. We generated a model with At-issueness (Prejacent vs. Belief) as a fixed effect and with random by-participant and by-item intercepts. We compared this model to a model without the At-issueness factor and found a significant effect ($\chi^2(1) = 11, p < 0.001$). As Figure 3 shows, sentences with at-issue prejacents were judged as more natural than sentences in which the belief component was at-issue. This is in line with the predictions of SSB, but not WB.

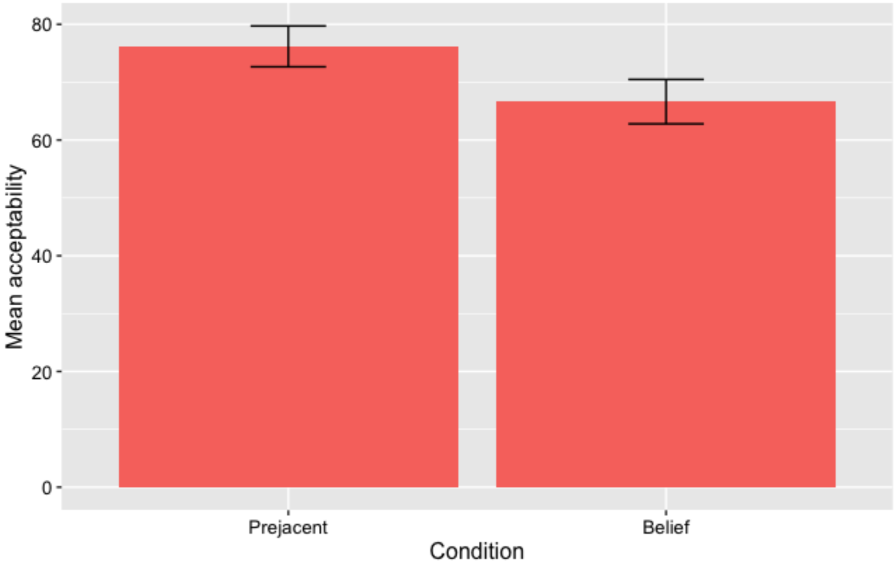


Figure 3: Average proportion of sentence acceptability ratings per condition in Experiment 1. Error bars indicate standard errors.

7. General discussion

The goal of this study was to test the divergent predictions of WB and SSB. We did this by conducting two experiments, which explored the predictions of these accounts as they related to grammatical factors (Experiment 1) and discourse factors (Experiment 2).

In Experiment 1 we varied the Person, Tense, and Position features of our test sentences. SSB predicted that sentences with first person/present tense/main clause features would be the most acceptable and that any changes toward a third person/past tense/embedded clause form would result in some reduction in acceptability. This is because, according to SSB, hedging uses of *believe* draw a contrast between private beliefs and public commitments, and so the closer the parameters on the measures Cr and Pr are to the utterance context parameters, the more salient and natural said contrast is predicted to be. In turn, WB is based on a single measure μ and predicted no substantive differences in acceptability between these different forms. Experiment 1 found evidence that, as SSB predicted, sentences in the past tense were less acceptable than those in the present tense and that embedded clauses were less acceptable than main clauses. This can be explained by pointing out that – when relativized to a past time t and a hypothetical world w – the measures Cr and Pr are detached from the utterance context and thus are less likely to be making claims about current probabilities. However, SSB's predictions were not supported by the results obtained for Person, as we found that sentences in the third person were overall more acceptable than those in the first person. We would like to make two points regarding this part of the results.

First, one possible reason why participants judged third person forms to be so natural might have been that third person belief reports naturally evoke a secondary speech context. This is due to the fact that people have no direct access to other people's mental states, so one's words are often reported as one's beliefs. Thus, *John believes it's raining* will typically be taken to imply *John said it's raining*, where the parameters on the belief measure *Cr* will correspond to $x =$ John, $w =$ the world of John's utterance, $t =$ the time of John's utterance.[14] If third person sentences are interpreted in this way, then it should be possible for participants to draw a subjective/objective contrast that is "shifted" to that secondary context. That is, SSB predicts that sentences like these are interpreted in a similar manner to the canonical form of first person/present tense/main clause and so acceptability is substantially improved. This line of explanation raises the question of why third person sentences were not merely rated similarly to first person sentences, but were actually rated (on average) as more acceptable. We suggest this is because first person forms cannot undergo a shift to a secondary speech context of the kind described above, the reason being that first person belief reports are based on direct experience and need not make reference to a prior utterance. But in the presence of past tense morphology or a hypothetical verb, the link of a first person belief sentence to the utterance context is easily lost and a hedging use quickly becomes implausible. The shifting mechanism thus makes third person forms more amenable to hedging uses overall.

Our second point regarding this result is that, while it was not predicted by the SSB, it was also not predicted by the WB, which did not expect any difference between first and third person forms. Therefore, we interpret this result as neutral with regard to these two accounts. Overall, we take the results of Experiment 1 as being more in line with the expectations of SSB than those of WB.

In Experiment 2 we explored the predictions of WB and SSB as they related to the acceptability of a hedging sentence in different discourse contexts. These contexts varied with regard to whether an (overt) QUD made the prejacent or the belief component at-issue. SSB predicted that such sentences should be judged as less natural when presented in contexts where the belief component is at-issue compared with contexts where the prejacent is at-issue, while WB does not distinguish between these contexts and so expects no difference in the level of sentence acceptability. The results of this experiment showed a significant difference in the acceptability of the test sentences in these two discourse contexts in the direction expected by SSB. That is, sentences in contexts where the prejacent was at-issue were rated as more natural than sentences in contexts where the belief component was at-issue. Therefore, we take the results of Experiment 2 as providing further empirical support for SSB over WB.

8. Outlook

This paper, in conjunction with Koev (2019), has established the plausibility of treating *believe* as a subjective modal and contrasting it with objective modals like *sure*. While our focus was on a specific lexical item, i.e. *believe*, the bigger question is how far the claimed distinction cuts into the epistemic domain more generally. Although we are not in position to make a concrete proposal at this point, we do have a tentative suggestion. We suggest that all epistemic modals are lexically encoded as subjective, objective, or either. Examples of the first category are *believe*, *think*, and *doubt*; an example of the second category is *sure*. The third, "neutral" category seems to encompass the largest class and includes things like *might*, *must*, *possible*, *likely*, *cer-*

[14] There are various ways to obtain such enriched interpretations. These include (i) assuming a richer semantics for *believe*, (ii) inserting a silent SAY operator on top of the structure, or (iii) assuming some form of a discourse-level mechanism (such as anaphora to a prior speech context or accommodation of such a context). We do not stake out a position here.

tain, etc. How does their neutrality come about? We do not believe that such neutral modals can be used subjectively or objectively because they are listed as ambiguous in the lexicon. Rather, we think that such modals are simply underspecified with regard to this dimension, and that the interpretation a given use receives depends on the context. One way to formally capture this is to assume that the semantics of such modals contains a free parameter that can be specified as either subjective or objective. We leave the substantiation of these rather speculative remarks to future work.

Appendix A

Test sentences in the first person + present tense + main clause condition (Experiment 1).

1. I believe it's raining, but I'm not sure.
2. I believe the Broncos will win, but I'm not sure.
3. I believe diplomacy is better than war, but I'm not sure.
4. I believe global warming is real, but I'm not sure.
5. I believe there's a fly in my office, but I'm not sure.
6. I believe Tiger Woods won 20 titles, but I'm not sure.
7. I believe medical cannabis is legal in all states, but I'm not sure.
8. I believe Mars has water, but I'm not sure.
9. I believe Friends is among the top ten American series, but I'm not sure.
10. I believe Florida has a warmer climate than Hawaii, but I'm not sure.
11. I believe children are getting lazier as technology progresses, but I'm not sure.
12. I believe 1962-1963 was the coldest winter on record, but I'm not sure.
13. I believe there are still enough natural resources on earth, but I'm not sure.
14. I believe there are more than 200 countries in the UN, but I'm not sure.
15. I believe Titanic won the most Oscars, but I'm not sure.
16. I believe Skechers has the best running shoes, but I'm not sure.

Appendix B

Trial test sentences in both the prejacent and the belief condition (Experiment 2).

1. A: Are Nike Zoom the best running shoes? (Prejacent)
   B: I believe they are, but I'm not sure.

   A: Why did you buy Nike Zoom running shoes? (Belief)
   B: I believe they are the best, but I'm not sure.

2. A: Will the Broncos win the game? (Prejacent)
   B: I believe they will, but I'm not sure.

   A: Why did you bet on the Broncos? (Belief)
   B: I believe they will win the game, but I'm not sure.

3. A: Is vaping safer than smoking? (Prejacent)
   B: I believe it is, but I'm not sure.

   A: Why did you stop smoking and start vaping? (Belief)
   B: I believe vaping is safer, but I'm not sure.

4. A: Is capitalism better than socialism? (Prejacent)
   B: I believe it is, but I'm not sure.

   A: Why do you want to preserve the free market? (Belief)
   B: I believe capitalism is better than socialism, but I'm not sure.

References

Barr, D. J., R. Levy, C. Scheepers, and H. J. Tily (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language 68(3), 255–278.

Bartsch, R. (1973). 'Negative transportation' gibt es nicht. Linguistische Berichte 27, 1–7.

Barwise, J. and R. Cooper (1981). Generalized quantifiers and natural language. Linguistics and Philosophy 4, 159–219.

Chemla, E. (2008). An epistemic step for anti-presuppositions. Journal of Semantics 25, 141–173.

Cresswell, M. J. (1976). The semantics of degree. In B. Partee (Ed.), Montague Grammar, pp. 261–292. New York: Academic Press.

Gajewski, J. R. (2007). Neg-raising and polarity. Linguistics and Philosophy 30, 289–328.

Hackl, M. (2009). On the grammar and processing of proportional quantifiers: *most* versus *more than half*. Natural Language Semantics 17, 63–98.

Hawthorne, J., D. Rothschild, and L. Spectre (2016). Belief is weak. Philosophical Studies 173, 1393–1404.

Hintikka, J. (1969). Semantics for propositional attitudes. In J. W. D. et al. (Ed.), Philosophical Logic, pp. 21–45. Dordrecht: Reidel.

Homer, V. (2015). Neg-raising and positive polarity: The view from modals. Semantics & Pragmatics 8(4), 1–88.

Horn, L. R. (1989). A Natural History of Negation. Chicago: University of Chicago Press.

Kennedy, C. (1999). Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison. Garland Press: New York.

Kennedy, C. and L. McNally (2005). Scale structure, degree modification, and the semantics of gradable predicates. Language 81(2), 345–381.

Koev, T. (2019). Strong beliefs, weak commitments. In Proceedings of Sinn und Bedeutung 23, vol. 2, pp. 1–18.

Kratzer, A. (1981). The notional category of modality. In H. J. Eikmeyer and H. Rieser (Eds.), Words, Worlds, and Contexts. New Approaches in Word Semantics, pp. 38–74. Berlin/New York: Walter de Gruyter.

Lassiter, D. (2017). Graded Modality: Qualitative and Quantitative Perspectives. Oxford: Oxford University Press.

Lyons, J. (1977). Semantics, vol. 2. Cambridge: Cambridge University Press.

Moore, G. E. (1993). Moore's paradox. In T. Baldwin (Ed.), G. E. Moore: Selected Writings, pp. 207–212. London: Routledge.

Nuyts, J. (2001). Subjectivity as an evidential dimension in epistemic modal expressions. Journal of Pragmatics 33, 383–400.

Papafragou, A. (2006). Epistemic modality and truth conditions. Lingua 116, 1688–1702.

Percus, O. (2006). Antipresuppositions. In A. Ueyama (Ed.), Theoretical and Empirical Studies of Reference and Anaphora: Toward the establishment of generative grammar as an empirical science, pp. 52–73. Report of the Grant-in-Aid for Scientific Research (B), Project No. 15320052, Japan Society for the Promotion of Science.

Portner, P. (2009). Modality. Oxford: Oxford University Press.

R Core Team (2020). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Romoli, J. (2013). A scalar implicature-based approach to neg-raising. Linguistics and Philosophy 36, 291–353.

Sauerland, U. (2008). Implicated presuppositions. In A. Steube (Ed.), The Discourse Potential of Underspecified Structures, pp. 581–600. Berlin: Mouton de Gruyter.

Schlenker, P. (2012). *Maximize Presupposition* and Gricean reasoning. Natural Language Semantics 20, 391–429.

Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. Lingua 117, 1034–1056.

Solt, S. (2016). On measurement and quantification: the case of *most* and *more than half*. Language 92(1), 65–100.

Unger, P. (1971). A defense of skepticism. The Philosophical Review 80(2), 198–219.

von Stechow, A. (1984). Comparing semantic theories of comparison. Journal of Semantics 3, 1–77.

Yalcin, S. (2010). Probability operators. Philosophy Compass 5(10), 916–937.