Natural Language Semantics manuscript No.
(will be inserted by the editor)

Modals under epistemic tensions:

Experimental evidence for the restricted quantificational account of *must* and *might*

Guillermo Del Pinal · Brandon Waldon

Forthcoming in *Natural Language Semantics*

Abstract According to Kratzer's influential account (1981; 1991; 2012), epistemic *must* and *might* involve quantification over domains of possibilities determined by a modal base and an ordering source. Recently, this account has been challenged by invoking contexts of 'epistemic tension': i.e., cases in which an assertion that *must* ϕ is conjoined with the possibility that $\neg\phi$, and cases in which speakers try to downplay a previous assertion that *must* ϕ, after finding out that $\neg\phi$. Epistemic tensions have been invoked from two directions. von Fintel and Gillies (2010) propose a return to a simpler modal logic-inspired account: *must* and *might* still involve universal and existential quantification, but the domains of possibilities are determined solely by realistic modal bases. In contrast, Lassiter (2016), following Swanson (2006, 2011), proposes a more revisionary account which treats *must* and *might* as probabilistic operators. In this paper, we present a series of experiments to obtain reliable data on the degree of acceptability of different contexts of epistemic tensions. Our experiments include novel variations that, we argue, are required to make progress in this debate. We show that restricted quantificational accounts *à la Kratzer* (e.g., Kratzer, 1981, 2012; Roberts, 2015; Giannakidou and Mari, 2016) fit the overall pattern of results better than either of their recent competitors. In addition, our results help us identify which components of restricted quantificational accounts are crucial for their success, and on that basis propose some general constraints that should be satisfied by all candidate accounts of the modal auxiliaries.

Keywords epistemic modals · epistemic tensions · modal reasoning · probabilistic reasoning · epistemic contradictions

This research was supported by grants from the Weinberg Institute for Cognitive Science at University of Michigan, Ann Arbor, Stanford University, and the Leibniz Center for General Linguistics, Berlin.

Guillermo Del Pinal
University of Michigan, Ann Arbor
E-mail: delpinal@umich.edu

Brandon Waldon
Stanford University
E-mail: bwaldon@stanford.edu

1 Introduction

1.1 Epistemic tensions and epistemic modals

Given information about the world, we can represent which states of affairs are possible, which are likely, and which are, though not directly known, still perfectly certain. How do we use such modal information to make inferences and decisions? Investigating the semantics of epistemic modals such as the auxiliaries *must* and *might*, and related adjectives like *certain* and *possible*, promises to shed light on this foundational question. In this paper, we present a series of experiments which use contexts of 'epistemic tension' to examine three influential accounts of the semantic structure and logical strength of these kinds of epistemic terms.

One way to introduce the competing accounts is to describe them in relation to the 'pure epistemic language hypothesis'. Suppose that, given states of knowledge, we can form representations of the set of worlds compatible with those states. Such epistemic spaces, '$\mathcal{E}$' for short, contain all and only those possibilities compatible with particular states of knowledge—nothing else constrains, structures, or orders $\mathcal{E}$:

Pure epistemic language hypothesis. The epistemic terms of natural languages are directly sensitive to $\mathcal{E}$. Theoretically, we can describe their semantics as operators over $\mathcal{E}$.

This hypothesis can be adopted in its full generality, or relative to a certain class of epistemic terms. It can also be categorically rejected—e.g., by arguing that all epistemic terms are directly sensitive to spaces that are more structured or constrained than $\mathcal{E}$. Two of the accounts we consider accept the pure epistemic language hypothesis with respect to *must* and *might*, and are compatible with, but do not directly defend, various extensions of it.

Focusing for concreteness on how epistemic *must* could operate on $\mathcal{E}$, consider the following view. An assertion of *it must be raining outside* is true if it is raining at the relevant location in *all* the worlds in $\mathcal{E}$. On this kind of account, recently defended by von Fintel and Gillies (2010), *must* is a maximally strong epistemic operator. This is prima facie plausible. Suppose Sam knows that her ball is in box A or B, and finds out that it isn't in A. Sam can then assert *the ball must be in B*, thereby expressing full certainty that the ball is in B. Consider, however, a context of 'epistemic tension' such as (1):

(1) Sara (addressing Jane): I just checked google and the weather channel. It must be raining outside. But if you insist, you can still go out and check, since we don't know for sure / it's strictly possible that it isn't...

Asserting that *it must be raining outside* seems intuitively appropriate in the context of (1). Yet this context includes the explicit admission that it is strictly possible that it isn't raining outside. In general, it seems that speakers produce *must*-claims in contexts where some epistemic doubt is acknowledged (Giannakidou and Mari, 2016; Lassiter, 2016, 2017). Contexts of epistemic tension such as (1) present a challenge to the strong account of *must*.

One reaction to tensions like (1) is to weaken the meaning of *must* while still holding on to the pure epistemic language hypothesis according to which it operates directly over $\mathcal{E}$. Here's one option: An assertion of *it must be raining outside*

is true if the proposition that it is raining has a high (but not necessarily maximal) probability in $\mathcal{E}$. This 'weak probabilistic' account is a simplified version of the account recently defended by Lassiter (2016), which in turn builds on Swanson (2006, 2011). This account can explain why, in (1), Sara can coherently assert that it must be raining outside and still admit the bare possibility of a world in $\mathcal{E}$ where it isn't raining at the relevant location. Consider, however, another kind of epistemic tension, the downplaying scenario in (2), due to von Fintel and Gillies (2010, 2016):

(2) a. Sara: I just checked google and the weather channel. It must be raining outside.
Jane opens the curtains. Sara and Jane both see that it is clearly not raining outside.
b. Jane: No it isn't. You were wrong!
c. Sara: # Well, strictly speaking, I was not wrong, I only said that it must be raining outside...

On the weak probabilistic view, the assertion that *it must be raining outside* in (2-a) is perfectly compatible with the possibility that, given what is known, there is a small chance that it isn't raining outside. Why, then, does Sara's final downplaying response, in (2-c), feel odd and unjustified? Given the hypothesized weakness of *must*, Sara's move should be perfectly adequate. Intuitively, however, it is not. In contrast, the strong view—according to which *must* expresses certainty given the evidence—can easily explain why Sara's reaction in (2-c) feels odd or unjustified (see von Fintel and Gillies, 2010, 2016).

At this point, we can see why contexts of epistemic tension have been used to test different theories of the meaning of epistemic terms (von Fintel and Gillies, 2010, 2016; Roberts, 2015; Giannakidou and Mari, 2016; Lassiter, 2016, 2017). Furthermore, tensions such as (1) and (2) seem to pose a real dilemma for two recent accounts, each compatible with the pure epistemic language hypothesis. One natural reaction to these cases is to try to explain the challenging data via additional general principles. Take the strong account of *must*. The intuitive acceptability of tensions like (1) could be due to their being interpreted as if they involve a mid-discourse change of mind or recognition that there are more possibilities than initially thought; or maybe such tensions feel acceptable because speakers often use *must* with pragmatic slack or license (cf. von Fintel and Gillies, 2010, 2016). It is hard to evaluate these sorts of responses based on our intuitions about isolated cases like (1). Indeed, as we will see, the intuitions of experts concerning even isolated tensions like (1) and (2) often diverge substantially.

One of the aims of our experiments, presented in §2 and §3, is to obtain reliable and systematic acceptability judgments for epistemic tensions of the form of (1) and (2). Specifically, we place contexts of epistemic tension in settings that allow us to examine whether versions of either the strong or the weak probabilistic accounts, when paired with additional general principles, predict the observed acceptability patterns. To do this, we obtained *comparative* acceptability ratings for tensions involving a wider range of epistemic terms than is usually considered, including various combinations of tensions with *must*, *might* and the related adjectives *certain* and *possible*. The overall pattern of results is quite rich, and we hope will be useful to theorists of different persuasions. Still, we will argue that the

results seriously challenge accounts which accept the pure epistemic language hypothesis, such as the strong and (some versions of) the weak probabilistic accounts of *must*.

The dilemma posed by contexts of epistemic tension, we will argue, is best resolved by accounts which adopt three key assumptions which rest on the uncontroversial claim that agents like us can represent not only pure epistemic spaces, but also epistemic spaces constrained by various expectations about the way the world is. (i) Epistemic terms are sensitive to these tiers of epistemic spaces, i.e., to sets of better and worse possibilities within $\mathcal{E}$. (ii) The auxiliaries *must* and *might* operate by default on the best worlds of $\mathcal{E}$. (iii) The adjectives *certain* and *possible* operate on wider domains which approximate pure epistemic spaces. On this picture, the auxiliaries quantify over subsets of $\mathcal{E}$ that are structured so as to be useful for particular contexts of deliberation. As a result, *must* claims are usually compatible with far off conflicting possibilities, as in cases like (1), but cannot be easily retracted in practical deliberation, explaining cases like (2). Theses (i)-(ii) are in line with the restricted quantificational account, due to Kratzer (1981, 1991, 2012), and elaborated in Roberts (2015) and Giannakidou and Mari (2016), and (iii) is broadly compatible with it. Our defense of (i)-(iii), thus, strongly supports this kind of accounts; but in addition, it will also issue in some significant refinements.

1.2 Theoretical background

To motivate the design and interpretation of our experiments, we need to first describe in more detail the competing accounts of *must* and *might*.[1] We also need to spell out what each account entails concerning the inferential relations between *must*, *certain*, *might* and *possible*. We begin by introducing some basic terminology and background assumptions.

To model the semantics of epistemic terms, we use basic tools from modal logic and probability theory. Some reminders and conventions will be helpful. Starting with the modal logic framework, let f_e stand for a function which, given a world w, returns a set of propositions that represent what is known in w (i.e., f_e stands for an *e*pistemic conversational background). Assuming that propositions are sets of possible worlds, $\bigcap f_e(w)$ stands for the set of worlds compatible with what is known in w. Since these are epistemic possibilities, $\bigcap f_e(w)$ picks out a realistic modal base, i.e., $w \in \bigcap f_e(w)$. By relativizing interpretations to epistemic modals bases, we can spell out the semantics of epistemic terms as quantifiers over that space.

For the probabilistic framework, we will use an implementation in which interpretations are relativized to functions e from worlds to epistemic probability spaces (Yalcin, 2010). An epistemic space is a pair $\langle E, Pr\rangle$ of a set of worlds E and a function Pr. E is a subset of the space of possible worlds W which corresponds to a set of worlds epistemically accessible from the evaluation world.[2] Pr is a function which assigns to each subset of W a number in [0, 1] satisfying:

[1] We initially adopt simple and unhedged versions of these accounts. In the General Discussion, §4, we consider different versions and modifications of the basic accounts.

[2] For simplicity, we assume that the set of all possible worlds W is finite (Yalcin, 2010).

(i) $Pr(E) = 1$ and (ii) if p and q are disjoint, $Pr(p \cup q) = Pr(p) + Pr(q)$. By relativizing interpretations to epistemic probability spaces, we can spell out the semantics of epistemic terms using Pr.

When comparing probabilistic and modal-logic based accounts, keep in mind that, for each w, $E_{e(w)}$, i.e., the set of epistemically accessible worlds determined by $e(w)$, is equivalent to $\bigcap f_e(w)$, i.e., the modal base determined by the epistemic conversational background $f_e(w)$. Just as context determines whose knowledge is represented by $f_e(w)$, context also determines whose knowledge is represented by $e(w)$. In other words, a semantic entry spelled out in terms of $e(w)$ is not, as such, more or less subjective or speaker oriented (when used in otherwise matching contexts) than one spelled out in terms of $f_e(w)$.

Our main background assumptions concern the epistemic adjectives *certain* and *possible*, as used in configurations like *it's certain/possible that* ϕ. Most theorists agree that *certain* is a maximally strong epistemic operator, while *possible* is a weak or minimum standard one. Our results, we will see, are consistent with this view (see Santorio and Romoli 2017 and Lassiter 2016). Additional details are more controversial: e.g., some think that they involve quantification over an epistemic modal base, as in (3), while others argue that they are probabilistic, as in (4). We can remain neutral on this specific debate (see Klecha, 2012; Kratzer, 2012; Lassiter, 2017; Santorio and Romoli, 2017).

(3) Strong *certain* + weak *possible* (modal logic version) :

a. $[\![certain\ \phi]\!]^{w,f_e} = 1$ iff $\forall w' \in \bigcap f_e(w) : [\![\phi]\!]^{w',f} = 1$

b. $[\![possible\ \phi]\!]^{w,f_e} = 1$ iff $\exists w' \in \bigcap f_e(w) : [\![\phi]\!]^{w',f} = 1$

(4) Strong *certain* + weak *possible* (probabilistic version) :

a. $[\![certain\ \phi]\!]^{w,e} = 1$ iff $Pr_{e(w)}(\{w' : [\![\phi]\!]^{w',e} = 1\}) = 1$

b. $[\![possible\ \phi]\!]^{w,e} = 1$ iff $Pr_{e(w)}(\{w' : [\![\phi]\!]^{w',e} = 1\}) > 0$

Our assumption, for now, is just that *it's certain that* ϕ is maximally strong: it involves universal quantification over an epistemic modal base, as in (3-a), or probability 1 in an epistemic probability space, as in (4-a). And *it's possible that* ϕ is very weak: it involves existential quantification, as in (3-b), or probability > 0 in an epistemic probability space, as in (4-b).[3]

Against this background, we now introduce each of our competing accounts of *must* and *might*. According to von Fintel and Gillies (2010), *must* is a universal and *might* an existential quantifier over epistemic modal bases. On this account, spelled out in (5), *must* is 'strong' in the sense that *must* ϕ entails that ϕ holds at each of the epistemically accessible worlds. Since $\bigcap f_e(w)$ stands for a realistic modal base, *must* ϕ entails ϕ. In addition, since *must* is maximally strong, it also follows that *must* ϕ entails *certain* ϕ. Strictly, these entailments only hold in contexts where the respective presuppositions are satisfied. Accordingly, the relevant notion—here and in what follows—is that of Strawson-entailment (see von Fintel, 1999; Sharvit, 2017). Assuming the usual dualities, *might* receives a

[3] This account of *certain* and *possible* is intended to cover their adjectival use in configurations like *it is certain/possible that* ϕ. We do not assume that it can be extended to the adverbs *certainly/possibly*. Also, note that, in §4, we will discuss some ways in which these entries for *certain* and *possible* can be relaxed, consistent with our experimental results.

weak semantics akin to something like bare possibility. As a result, *possible* ϕ entails *might* ϕ.[4]

(5) Strong *must* + weak *might*:

a. $[\![must\ \phi]\!]^{w,f_e} = 1$ iff $\forall w' \in \bigcap f_e(w) : [\![\phi]\!]^{w',f} = 1$

b. $[\![might\ \phi]\!]^{w,f_e} = 1$ iff $\exists w' \in \bigcap f_e(w) : [\![\phi]\!]^{w',f} = 1$

c. S-entailments:

(i) *must* $\phi \models \phi$

(ii) *must* $\phi \models$ *certain* ϕ

(iii) *possible* $\phi \models$ *might* ϕ

Recall that holding that *must* is strong helps explain why downplaying in cases like (2) feels odd, or unjustified. At the same time, this account is challenged by cases like (1) where *must* ϕ assertions occur in contexts where the possibility that $\neg\phi$ is explicitly acknowledged.[5]

Many semanticists hold, pace von Fintel and Gillies (2010), that epistemic *must* is weaker than strict necessity and *might* stronger than bare possibility. Yet these accounts can be divided into two different camps. The first camp adopts a probabilisitic framework, as originally proposed by Swanson (2006) and recently elaborated and defended by Lassiter (2016, 2017). We focus on a probabilistic implementation, presented in (6), that is consistent with the pure epistemic language hypothesis.[6] On this view, *must* ϕ entails that the probability of ϕ, $Pr_{e(w)}(\phi)$, is above some high but non-maximal contextually determined threshold θ_c, as in (6-a). It follows that *must* ϕ is compatible with the possibility that $\neg\phi$, as long as the probability of the latter is $< 1 - \theta_c$. In addition, by comparing (4-a) and (6-a) we can easily see that, on this account, *certain* ϕ asymmetrically entails *must* ϕ. To maintain the standard dualities, we can hold that *might* ϕ entails that $Pr_{e(w)}(\phi) > 1 - \theta_c$.[7]

(6) Weak *must* + strong *might* (probabilistic version):

a. $[\![must\ \phi]\!]^{w,e} = 1$ iff $Pr_{e(w)}(\{w' : [\![\phi]\!]^{w',e} = 1\}) > \theta_c$

b. $[\![might\ \phi]\!]^{w,e} = 1$ iff $Pr_{e(w)}(\{w' : [\![\phi]\!]^{w',e} = 1\}) > 1 - \theta_c$

c. S-entailments:

(i) *must* $\phi \not\models \phi$

(ii) *must* $\phi \not\models$ *certain* ϕ

(iii) *certain* $\phi \models$ *must* ϕ

(iv) *possible* $\phi \not\models$ *might* ϕ

(v) *might* $\phi \models$ *possible* ϕ

[4] Given entries like (3) for *certain* and *possible*, the entailments from *certain* to *must* and from *might* to *possible* also obtain. However, theorists who favor (5-a)-(5-b) could reject that *certain* is maximally strong and that *possible* is just bare possibility; at any rate, our arguments don't depend on these additional entailments.

[5] von Fintel and Gillies (2010) argue that the perceived weakness of *must* is due to an indirectness presupposition or implication. This can explain—without conceding weakness—the oddness of cases as when Jimmy, staring directly at the pouring rain, asserts *it must be raining outside*. Crucially, the stimuli used in all of our experiments were designed to satisfy this indirectness requirement, so we ignored this (potential) evidential component in the lexical entries for *must* and *might*.

[6] In the General Discussion, we discuss probabilistic accounts which abandon the pure epistemic language hypothesis.

[7] Lassiter's probabilistic account of *must*/*might* is more complex than (6). However, as we argue in §4 that, given our experimental stimuli/designs, Lassiter's account doesn't make different predictions than the simpler version in (6).

Recall that assuming that *must* is weak helps explain epistemic tensions, like (1), where felicitous uses of *must* co-exist with lack of full certainty in the prejacent. On the other hand, this account has difficulty explaining why we can't downplay previous *must*-claims in epistemic tensions like (2).[8]

Theorists in the second camp hold that, as in modal logic, *must* and *might* involve universal and existential quantification (Kratzer, 1991, 2012; Roberts, 2015; Giannakidou and Mari, 2016). However, following Kratzer (1991, 2012), they abandon the pure epistemic language hypothesis. The domain of quantification of *must* is determined by an epistemic modal base *and* an ordering source, as in (7-a). Ordering sources for epistemic modals are typically generated from a 'stereotypical conversational background', g_s: $g_s(w)$ returns a set of propositions representing what is taken to be normal in w. $g_s(w)$ is used to pick out a subset of the 'best' or most 'normal' worlds in $\bigcap f_e(w)$.[9] Unlike $\bigcap f_e(w)$, the subset of the best worlds, $\text{BEST}_{g_s(w)}(\bigcap f_e(w))$, is not realistic: in any situation of partial knowledge, we could be in a world that turns out be abnormal. As a result, *must* ϕ does not entail ϕ. Furthermore, since $\text{BEST}_{g_s(w)}(\bigcap f_e(w)) \subseteq \bigcap f_e(w)$, *must* is strictly weaker than *certain*, when the latter is defined as in (3-a). Given the standard dualities, *might* is now stronger than just bare possibility, as captured in (7-b).[10]

(7) Weak *must* + strong *might* (restricted quantification version):

a. $[\![must\ \phi]\!]^{w,f_e,g_s} = 1$ iff $\forall w' \in \text{BEST}_{g_s(w)}(\bigcap f_e(w)) : [\![\phi]\!]^{w',f,g} = 1$

b. $[\![might\ \phi]\!]^{w,f_e,g_s} = 1$ iff $\exists w' \in \text{BEST}_{g_s(w)}(\bigcap f_e(w)) : [\![\phi]\!]^{w',f,g} = 1$

c. S-entailments:

(i) *must* $\phi \not\models \phi$
(ii) *must* $\phi \not\models$ *certain* ϕ
(iii) *certain* $\phi \models$ *must* ϕ
(iv) *possible* $\phi \not\models$ *might* ϕ
(v) *might* $\phi \models$ *possible* ϕ

Based on the entailments in (6-c) and (7-c), one might conclude that the probabilistic and universal versions of the weak *must* + strong *might* account have the same virtues and problems relative to contexts of epistemic tension. Despite this tempting conclusion, we will argue in §3 that these accounts make different

[8] For the weak probabilistic view in (6) to really preserve the pure epistemic language hypothesis, we must put further constraints on Pr. Specifically, we must prevent Pr from incorporating non-epistemic information, such as normality assumptions. One option: for each context c, Pr assigns a nonzero probability to each world in E_c. In §4, we discuss which constrains on Pr would account for our results, and argue that the result is a probabilistic account that abandons the pure epistemic language hypothesis.

[9] We follow the usual simplifications and procedures for deriving the set of best worlds given a modal base $\bigcap f_e(w)$ and ordering source $g_s(w)$ (Portner, 2009; Hacquard, 2011). Following the procedure in (i), $g_s(w)$ induces an ordering $\preceq_{g_s(w)}$ on a set of worlds W: intuitively, w is at least as good as z if w makes true at least as many propositions in $g_s(w)$ as z does. Given the limit assumption, we can then use $\preceq_{g_s(w)}$ to pick out the best worlds, as in (ii).

(i) $\forall w, z \in W : w \preceq_{g_s(w)} z$ iff $\{p : p \in g_s(w) \text{ and } z \in p\} \subseteq \{p : p \in g_s(w) \text{ and } w \in p\}$

(ii) $\text{BEST}_{g_s(w)}(\bigcap f(w)) := \left\{v : v \in \bigcap f(w)\ \&\ \neg\exists v' \in \bigcap f(w) : v' \prec_{g_s(w)} v\right\}$

[10] There are important differences between various versions of the restricted quantificational account, some of which we discuss in §4.3. For example, Giannakidou and Mari (2016) hold that *might* is a bare possibility operator, although their account can be easily revised to make this entry stronger. It is also worth pointing out that, although we won't discuss dynamic accounts here, some recent proposals follow the spirit of the account in (7). One example is Willer (2013), who also models *might* as, roughly, live or serious possibility.

predictions in epistemic tensions like (2), and that the restricted quantificational account makes better predictions.

2 Experiments 1A-B

Experiments 1A-B test pit the strong *must* + weak *might* package against the two weak *must* + strong *might* packages by examining the acceptability ratings of expressions in which pairs of epistemic clauses are combined so as to generate varying degrees of epistemic tension, as in (1) above. We said at the outset that cases like (1) favor packages with weak accounts of *must*. However, to appreciate the rationale for our design, we should now note that such cases have also been invoked to support accounts of strong *must*.

In particular, von Fintel and Gillies (2010) report that the examples in (8) feel contradictory, and they argue that this judgment is hard to explain for packages that accept weak *must*. For if weak, *must* ϕ would be strictly compatible with the possibility that $\neg\phi$, and cases like (8) should have acceptable readings. von Fintel and Gillies (2010) predict that even if *perhaps* itself is given a non-weak semantics, minimal variants obtained by replacing *perhaps* with any other weak possibility modal still feel contradictory.

(8) a. #It must be raining, but perhaps it isn't.
b. #Perhaps it isn't raining, but it must be.

However, as illustrated in (1), placing the expressions of epistemic tension in supporting contexts seems to improve their acceptability, and so does using minimal variants with other weak epistemic operators (Giannakidou and Mari, 2016; Lassiter, 2016; Roberts, 2015). Consider (9):

(9) . . . I think we can just forget about today's climb. We just saw the weather forecast, and
a. ?it must be raining; although it's strictly possible that it isn't.
b. ?although it's possible that it isn't, it must be raining.

What can we conclude from this impasse? Since the competing accounts make unique predictions in cases like (1), (8) and (9), examining the degree of acceptability of expressions of epistemic tensions remains a potentially useful source of evidence for debates about the strength of modals. Still, it's undeniable that the target judgments are subtle. To properly use this strategy, we need to obtain the acceptability judgments via systematic empirical methods.

Indeed, empirical work is beginning to play a key role in this debate. In particular, Lassiter (2016) presents corpus data showing that expressions akin to (1) and (9) are commonly used by ordinary speakers. He argues that this undermines the reliability of the intuitions reported by von Fintel and Gillies (2010) with respect to cases like (8) and (9), and supports the view that *must* is weak. To reconcile these general results with (8), Lassiter suggests, *contra* von Fintel and Gillies (2010), that terms like *perhaps* and *might* are stronger than genuine bare possibility modals and operators.

Lassiter's data leaves open several key issues, however, some connected with more general limitations of corpus data:

(i) Though suggestive, the fact that expressions of epistemic tension like (8) and (9) can be found in corpus data does not show that they aren't odd (cf. von Fintel and Gillies, 2016). After all, we presumably sometimes produce semantically deficient expressions.
(ii) There is empirical work suggesting that, for ordinary speakers, some contradictions are acceptable (Sauerland, 2011; Alxatib et al, 2013). What should we conclude if, despite being naturalistically produced, expressions like (8) and (9) are no more acceptable than some straightforward contradictions?
(iii) Epistemic tensions using *must* and *might* do not seem to have been generated nearly as frequently as examples akin to (8) and (9), which involve other expressions of weak epistemic possibility. Does this suggest that *must* and *might* are indeed duals? Does it suggest that *might* is stronger than other expressions of bare epistemic possibility?
(iv) Corpus data can't be used to directly address questions about the comparative strength of different epistemic operators, yet this information is crucial to determine the strength of *must* and *might*. So we need to directly examine possible significant differences in the relative acceptability of various expressions of epistemic tension.

Issues (i)-(iv) suggest that obtaining corpus data on the use of expressions of epistemic tension is only a first step in the use of empirical methods to inform debates about the strength of epistemic modals. We also need to obtain ratings that can be used to determine the comparative acceptability of different expressions of epistemic tension. This is the goal of Experiments 1A-B. Specifically, we compare the acceptability of minimal variants involving epistemic tensions such as: *must* ϕ $\wedge$ *possible* $\neg\phi$, *must* ϕ $\wedge$ *might* $\neg\phi$, *must* ϕ $\wedge$ $\neg\phi$, and also *certain* ϕ $\wedge$ *possible* $\neg\phi$, *certain* ϕ $\wedge$ *might* $\neg\phi$, *certain* ϕ $\wedge$ $\neg\phi$. Since the main competing accounts make different predictions in some of these cases, the results can directly bear on whether we should adopt the strong *must* + weak *might* package, or one of the weak *must* + strong *might* packages.

2.1 Experiment 1A

2.1.1 Materials and Methods

Experiment 1A consisted of a between-subjects design where each participant rated the acceptability of an epistemic tension, on a scale from 1 ('completely unacceptable') to 7 ('completely acceptable'), presented in bold within a 3-4 sentence vignette. We based our stimuli on the corpus data obtained by Lassiter (2016), where *must* ϕ statements occur with explicit avowals of the possibility that $\neg\phi$ via statements such as *it is possible that not* ϕ and *I don't know for sure that* ϕ. When appropriate, we modified the original passages to enhance coherence. We constructed five distinct discourse passages. A representative passage is (10). The target sentence in each passage was taken from four conditions, as in (10-a)-(10-d). The conditions in (10-a) and (10-b)—namely, *must* ϕ $\wedge$ *not sure* ϕ and *must* ϕ $\wedge$ *possible* $\neg\phi$—consist of combinations of *must* ϕ with avowals of the bare possibility that $\neg\phi$. The condition in (10-c) combines *must* ϕ with claims of the form *might* $\neg\phi$. And the condition in (10-d) captures the strongest epistemic tension, where *must* ϕ is simply combined with the assertion that $\neg\phi$.

(10) I just bought a vintage bicycle at a garage sale in my neighborhood. It will need some work, but it's in decent shape. [EPISTEMIC TENSION HERE] The previous owner didn't know the name of the manufacturer.

a. 'must $\phi \wedge$ not sure ϕ':
The bike must be from the 60s, but I don't know for sure.

b. 'must $\phi \wedge$ possible $\neg\phi$':
The bike must be from the 60s, but it's possible that it isn't.

c. 'must $\phi \wedge$ might $\neg\phi$':
The bike must be from the 60s, but it might not be.

d. 'must $\phi \wedge \neg\phi$':
The bike must be from the 60s, but it isn't.

Crucially, we presented each of our target sentences in two orders. In one order the *must*-statement comes first, as in (10-a)-(10-d). In the reverse order the *must*-statement appeared last, as in *it is possible that the bike isn't from the 60s, but it must be*. Summing up, there were 4 possible combinations for each *must* ϕ statement, each presented in 2 distinct orders. This results in a total of 8 target conditions across the 5 discourse contexts. Each of the resulting 40 critical sentences was evaluated by 40 unique participants.

The materials presented so far were designed to compare the degree of acceptability of tensions which combine *must* ϕ and the bare possibility that $\neg\phi$, with tensions which combine *must* ϕ and *might* $\neg\phi$. However, we also wanted to compare each of the epistemic tensions involving *must*-statements with minimal variants involving uncontroversially strong epistemic operators. To do so, we included matching conditions obtained as follows: the *must* ϕ conjunct in each condition was replaced with a sentences of the form *it's certain that* ϕ. The matching passage for (10) is presented in (11), and the matching conditions for (10-a)-(10-d) are presented in (11-a)-(11-d):

(11) I just bought a vintage bicycle at a garage sale in my neighborhood. It will need some work, but it's in decent shape. [EPISTEMIC TENSION HERE] The previous owner didn't know the name of the manufacturer.

a. 'certain' $\phi \wedge$ not sure ϕ':
It's certain that the bike is from the 60s, but I don't know for sure.

b. 'certain $\phi \wedge$ possible $\neg\phi$':
It's certain that the bike is from the 60s, but it's possible that it isn't.

c. 'certain $\phi \wedge$ might $\neg\phi$':
It's certain that the bike is from the 60s, but it might not be.

d. 'certain $\phi \wedge \neg\phi$':
It's certain that the bike is from the 60s, but it isn't.

As before, we presented each of these target sentences in two orders, one illustrated in (11-a)-(11-c) and the other in the reverse order of the conjuncts. As a result, we again had 40 critical sentences, and each was evaluated by 40 participants. (The full set of stimuli used in Experiment 1A is included in Appendix A.)

We recruited 3,200 participants using Amazon Mechanical Turk (AMT). We recruited only AMT workers with US internet protocol addresses who possessed an approval rating greater than 90% for past tasks completed on AMT. The experiment was deployed across several AMT assignments (HITs), which were all

simultaneously published on July 17th, 2017 around 11:30 am PDT, and concluded on July 28, 2017 at around 1:30 AM PDT. To ensure that workers did not complete our experiment more than once, we used the UniqueTurker script (https://uniqueturker.myleott.com/), which restricted workers to completing just one of our HITs. Workers were also notified that "There are many versions of this experiment currently online. Please only complete ONE version; you will not be paid for participating more than once". These two safeguards were mostly successful: two participants tried to do the experiment more than once, and were thus excluded from the analysis. We did, however, allow participation from workers who may have seen a similar pilot version of this experiment roughly seven weeks prior.

2.1.2 Predictions

Experiment 1A was designed to test the strong *must* + weak *might* package vs. the weak *must* + strong *might* packages. It was not designed to distinguish between either version of the weak *must* + strong *might* package. Before spelling out the predictions of these accounts, three clarifications are in order. First, recall our assumptions about the epistemic adjectives, laid out in (3)-(4): *certain*, when used as in *it is certain that* ϕ, is a maximally strong epistemic operator, and *possible*, when used as in *it is possible that* ϕ, is a weak possibility operator. It follows that, in our target configurations, *possible* has the same logical strength as the weak *might* entry in (5-b), and *certain* has the same strength as the strong *must* entry in (5-a). Second, we spell out the predictions of these accounts as insensitive to the order in which the conjuncts are presented. We discuss the possible relevance of order in §2.1.3. Third, when stating the predictions below, we use 'weakly possible $\neg\phi$' to stand for both the 'not sure ϕ' conjunct and the 'possible $\neg\phi$' conjunct.

Consider first the unique predictions of the strong *must* + weak *might* account, summarized in (12) (where '$x \succ y$' means that 'x is more acceptable than y'). On this view, spelled out in (5), *must* ϕ entails ϕ, and also that ϕ is certain. In addition, *might* ϕ receives a very weak interpretation, akin to the bare possibility that ϕ. Accordingly, this view predicts that each of the 'must $\phi \wedge$ weakly possible $\neg\phi$' conditions should be approximately as unacceptable as the 'must $\phi \wedge$ might $\neg\phi$' condition. The reason is that these conditions all involve assertions which entail ϕ, conjoined with assertions that it is possible that $\neg\phi$. This prediction is captured in (12-a). What are the predictions for the pairwise comparisons of epistemic tensions involving *must* and the matching variants with *certain*? Since on this view *must* ϕ entails that ϕ is certain, it follows that an expression of epistemic tension which conjoins *must* ϕ with the possibility that $\neg\phi$ should be at least as unacceptable as one which conjoins *certain* ϕ with the possibility that $\neg\phi$. This prediction is captured in (12-b).

(12) Strong *must* + weak *might*:
   a. 'must $\phi \wedge$ weakly possible $\neg\phi$' $\approx$ 'must $\phi \wedge$ might $\neg\phi$'
   b. 'must $\phi \wedge$ weakly possible $\neg\phi$' $\preceq$ 'certain $\phi \wedge$ weakly possible $\neg\phi$'

Consider next the unique predictions of the weak *must* + strong *might* packages, summarized in (13). On this view, *must* ϕ is compatible with the bare possibility that $\neg\phi$. At the same time, *must* ϕ is not compatible with *might* $\neg\phi$. This

is because while weak possibilities existentially quantify over the whole epistemic modal base, *must* and *might* quantify only over a subset—i.e., the best worlds—of the epistemic modal base. As a result, this view predicts that each of the 'must $\phi \wedge$ weakly possible $\neg\phi$' conditions should be more acceptable than the 'must $\phi \wedge$ might $\neg\phi$' condition. This prediction is captured in (13-a). What about the predictions for the pairwise comparisons of epistemic tensions involving *must* and the matching variants with *certain*? On this view, *must* ϕ does *not* entail that ϕ is certain: the former is compatible, while the latter incompatible, with the bare possibility that $\neg\phi$. Accordingly, an expression of epistemic tension which conjoins *must* ϕ with the possibility that $\neg\phi$ is predicted to be substantially more acceptable than one which conjoins *certain* ϕ with the possibility that $\neg\phi$. This prediction is captured in (13-b).

(13) Weak *must* + strong *might* packages:
   a. 'must $\phi \wedge$ weakly possible $\neg\phi$' $\succ$ 'must $\phi \wedge$ might $\neg\phi$'
   b. 'must $\phi \wedge$ weakly possible $\neg\phi$' $\succ$ 'certain $\phi \wedge$ weakly possible $\neg\phi$'

2.1.3 Results and discussion

Mean ratings across the 8 conditions involving the *must* ϕ statements are presented in Table 1 (recall that ratings were based on a 1 'completely unacceptable' to 7 'completely acceptable' scale). Data from 6 participants who left a blank response were excluded. Using the ordinal package in R (Christensen 2015), we conducted an ordinal mixed effects logistic regression predicting acceptability ratings from a fixed effect of discourse type with the plain contradiction 'must $\phi \wedge \neg\phi$' as the reference level for the conditions in the standard-order, and '$\neg\phi \wedge$ must ϕ' as the reference level for the conditions in the reverse order. In both cases, the model included by-item random intercepts and by-item slopes for condition. In the standard order, acceptability ratings were significantly higher for the 'must $\phi \wedge$ not sure ϕ' condition ($\beta = 1.41$, $SE = 0.243$, $z = 5.82$, $p < 0.0001$), as well as for the 'must $\phi \wedge$ possible $\neg\phi$ condition ($\beta = 0.554$, $SE = 0.178$, $z = 3.06$, $p < 0.003$). In contrast, the acceptability ratings were not significantly higher for the 'must $\phi \wedge$ might $\neg\phi$' condition ($\beta = -0.060$, $SE = 0.178$, $z = -0.336$, $p < 0.74$). The same pattern of results was observed in the reverse order. Acceptability ratings were significantly higher for the 'not sure $\phi \wedge$ must ϕ' condition ($\beta = 2.05$, $SE = 0.237$, $z = 8.69$, $p < 0.0001$), as well as for the 'possible $\neg\phi \wedge$ must ϕ' condition ($\beta = 1.03$, $SE = 0.186$, $z = 5.56$, $p < 0.0001$). In contrast, acceptability ratings were not higher for the 'might $\neg\ \phi \wedge$ must ϕ' condition ($\beta = -0.162$, $SE = 0.183$, $z = -0.891$, $p < 0.38$).

The basic pattern of results is simple. Each combination of *must* ϕ with the weak possibility that $\neg\phi$ was significantly more acceptable than the baseline epistemic tension/contradiction *must* $\phi \wedge \neg\phi$. In contrast, the *must* $\phi \wedge$ *might* $\neg\phi$ condition obtained very low ratings: specifically, the ratings were statistically indistinguishable from those obtained for the baseline epistemic contradiction. This pattern of results confirms prediction (13-a) of the weak *must* + strong *might* packages, and is in tension with prediction (12-a) of the strong *must* + weak *might* package. Specifically, the overall pattern supports the view that *might*-statements are stronger than matching bare possibility statements. Crucially, this

Standard order	**Mean Rating**	**Reverse order**	**Mean Rating**
must $\phi \wedge$ not sure ϕ	5.37 (SD = 1.62)	not sure $\phi \wedge$ must ϕ	5.26 (SD = 1.55)
must $\phi \wedge$ possible $\neg\phi$	4.62 (SD = 1.67)	possible $\neg\phi \wedge$ must ϕ	4.30 (SD = 1.78)
must $\phi \wedge$ might $\neg\phi$	4.01 (SD = 1.69)	might $\neg\phi \wedge$ must ϕ	3.16 (SD = 1.62)
must $\phi \wedge \neg\phi$	4.08 (SD = 1.76)	$\neg\phi \wedge$ must ϕ	3.33 (SD = 1.74)

Table 1 Average ratings for each epistemic tension condition which included 'must' in Experiment 1A. The key result was that although the ratings for epistemic tensions involving *might* were indistinguishable from the baseline contradictions, each of tension involving a weak possibility modal was rated as significantly more acceptable.

basic pattern of results is observed independently of the order of presentation of the conjuncts.

If one accepts that *might* is stronger than bare possibility, the results reported in Table 1 can still be squared with the view that *must* is strong. Proponents of this view could argue that the difference between, say, the *must* $\phi \wedge$ *possible* $\neg\phi$ condition and the *must* $\phi \wedge$ *might* $\neg\phi$ condition is just that the latter contradiction is 'stronger' or more 'salient' in some informal psychological sense reflected in the acceptability ratings. For this reason, to obtain direct evidence about the strength of *must* we need to also compare the scores of each *must* ϕ condition with its matching *certain* ϕ condition. If *must* is strong, then each *must* condition should be at least as degraded as its matching *certain* condition. The average ratings across the 8 conditions involving the *certain* ϕ statements are presented in Table 2. Pairwise ordinal mixed effects logistic regression analyses were conducted for each 'certain' condition against its matching 'must' condition. Each model included included by-item random intercepts and by-item slopes for condition.[11] The difference was significant for each pairwise comparison ($p < 0.0001$), with the exception of 'certain $\phi \wedge$ might $\neg\phi$' vs. 'must $\phi \wedge$ might $\neg\phi$', which is significant at $p < 0.04$, and 'might $\neg\phi \wedge$ certain ϕ' vs. 'might $\neg\phi \wedge$ must ϕ', which on this analysis was marginally significantly ($p < 0.08$).[12]

The basic pattern of results is again remarkably simple. As expected in this kind of task, the overall ratings reflect the non-formal, intuitive strength of the tensions, with the uncontroversial strongest baseline tensions/contradictions receiving the lowest score. What is crucial, however, is that each *certain* condition received a lower acceptability rating than its matching *must* condition. These results are unexpected if we assume that *must* is strong, and in particular that *must* ϕ entails that ϕ is certain. Specifically, the results disconfirm prediction (12-b) of the strong *must* package. In contrast, this pattern of results is expected on the assumption that *must* is weak, i.e., that assertions that *must* ϕ require less than certainty that

[11] Results for each pairwise comparison:
'certain $\phi \wedge$ not sure ϕ' vs. 'must $\phi \wedge$ not sure ϕ': $\beta = 1.20$, $SE = 0.201$, $z = 5.91$, $p < 0.0001$
'certain $\phi \wedge$ possible $\neg\phi$' vs. 'must $\phi \wedge$ possible $\neg\phi$': $\beta = 0.825$, $SE = 0.180$, $z = 4.59$, $p < 0.0001$
'certain $\phi \wedge$ might $\neg\phi$' vs. 'must $\phi \wedge$ might $\neg\phi$': $\beta = 0.371$, $SE = 0.178$, $z = 2.09$, $p < 0.04$
'certain $\phi \wedge \neg\phi$' vs. 'must $\phi \wedge \neg\phi$': $\beta = 0.763$, $SE = 0.180$, $z = 4.24$, $p < 0.0001$
'not sure $\phi \wedge$ certain ϕ' vs. 'not sure $\phi \wedge$ must ϕ': $\beta = 2.16$, $SE = 0.364$, $z = 5.93$, $p < 0.000$'
'possible $\neg\phi \wedge$ certain ϕ' vs. 'possible $\neg\phi \wedge$ must ϕ': $\beta = 1.30$, $SE = 0.212$, $z = 6.12$, $p < 0.0001$
'might $\neg\phi \wedge$ certain ϕ' vs. 'might $\neg\phi \wedge$ must ϕ': $\beta = 0.368$, $SE = 0.205$, $z = 1.79$, $p < 0.08$
'$\neg\phi \wedge$ certain ϕ' vs. '$\neg\phi \wedge$ must ϕ': $\beta = 1.10$, $SE = 0.217$, $z = 5.06, p < 0.0001$

[12] But note these last comparisons don't play an important role in our dialectic because our competing theories all make the same predictions in the tensions involving *might*, namely, they are all predicted to be contradictory, hence roughly as unacceptable as plain contradictions, which is basically what we observe.

Standard order	**Mean Rating**	**Reverse order**	**Mean Rating**
certain $\phi \wedge$ not sure ϕ	4.12 (SD = 1.96)	not sure $\phi \wedge$ certain ϕ	3.21 (SD = 1.84)
certain $\phi \wedge$ possible $\neg\phi$	3.76 (SD = 1.89)	possible $\neg\phi \wedge$ certain ϕ	3.01 (SD = 1.77)
certain $\phi \wedge$ might $\neg\phi$	3.65 (SD = 1.82)	might $\neg\phi \wedge$ certain ϕ	2.86 (SD = 1.63)
certain $\phi \wedge \neg\phi$	3.34 (SD = 1.69)	$\neg\phi \wedge$ certain ϕ	2.42 (SD = 1.61)

Table 2 Average ratings for each epistemic tension condition which included 'certain' in Experiment 1A.The key result was that each epistemic tension with *certain* ϕ was less acceptable than the matching tension with *must* ϕ.

ϕ. Specifically, the results confirm prediction (13-b) of the weak *must* package. As before, this basic pattern was preserved independently of the order in which the conjuncts were presented.

Overall, the results of Experiment 1, summarized in Tables 1-2, support the weak *must* + strong *might* packages, and are in tension with the strong *must* + weak *might* package. According to the former, *must* is weaker than *certain*, and *might* is stronger than *possible*, which corresponds to the observed acceptability patterns. Interestingly, these results are compatible with the view that *must*/*might* and *certain*/*possible* are duals. If this were the case, the combinations of these duals in expressions of epistemic tension should be contradictory and equally bad. To test this prediction, we performed pairwise comparisons for the following pairs of conditions: 'certain $\phi \wedge$ possible $\neg\phi$' vs. 'must $\phi \wedge$ might $\neg\phi$' and also in the reverse order, namely, 'possible $\neg\phi \wedge$ certain ϕ' vs. 'might $\neg\phi \wedge$ must ϕ'. The differences between these conditions were not significant in either order.[13]

Thus far, we considered the predictions of the competing accounts directly, as if there were no other factors. At this point, one might ask whether we can reconcile the results of Experiment 1A with the strong *must* + weak *might* package by appealing to independently motivated general principles. For example, after granting, in response to Lassiter's corpus data, the intuitive acceptability of epistemic tensions such as (10), von Fintel and Gillies (2016) stick to the strong *must* account and propose a solution which goes roughly as follows. Examples like (10) feel acceptable because they involve a dynamic change in the modal horizon or epistemic modal base. This dynamic change can result in expansion or contraction of the modal base, depending on the order of the conjuncts that are being accommodated. This is a reasonable way of trying to explain the perceived acceptability of instances of *must* $\phi \wedge$ *possible* $\neg\phi$ (for both the standard and reverse order of the conjuncts).

Evaluating this response on behalf of the strong *must* + weak *might* package illustrates the importance of having obtained, in Experiment 1A, comparative acceptability ratings. Specifically, although it is hard to evaluate this response by looking just at data on the *must* conditions, we can evaluate it by comparing the acceptability ratings of the *must* vs. *certain* conditions. Since this response appeals to a *general* principle of information dynamics and accommodation, it should also apply to the epistemic tensions in the matching *certain* conditions. If we can easily expand the modal horizon to accommodate an assertion of *possible* $\neg\phi$ after an assertion of *must* ϕ, we should also be able to do so to accommodate an assertion

[13] For 'certain $\phi \wedge$ possible $\neg\phi$' vs. 'must $\phi \wedge$ might $\neg\phi$', the results were as follows: $\beta = 0.264$, $SE = 0.178$, $z = 1.47$, $p < 0.15$. For the reverse order, namely, 'possible $\neg\phi \wedge$ certain ϕ' vs. 'might $\neg\phi \wedge$ must ϕ', the results were: $\beta = 0.247$, $SE = 0.257$, $z = 0.959$, $p < 0.34$.

of *possible* $\neg\phi$ after an assertion of *certain* ϕ—especially if we assume that, in this context, *must* and *certain* have roughly equal epistemic strength. However, what has to be explained is why the tensions involving *must* are significantly more acceptable than the matching tensions with *certain*. So conjoining the strong *must* account with this general principle does not help explain the key pattern of results observed in Experiment 1A: namely, that the tensions with *must* are significantly more acceptable than the tensions with *certain*. (It is easy to see that an analogous problem arises if this kind of response is used to defend the view that *might* expresses bare possibility).[14]

A similar response on behalf of the strong *must* package is to appeal to 'pragmatic slack' (Lasersohn, 1999). As before, this could explain why, on a strong view, epistemic tensions with *must* $\phi \wedge$ *possible* $\neg\phi$ can feel acceptable (see Klecha, 2014; von Fintel and Gillies, 2016). However, appealing to slack, by itself, doesn't account for the key contrast observed in Experiment 1. Suppose slack can resolve the target epistemic tensions when fed a strong *must*, so that *must* ϕ entails that ϕ and that ϕ is certain. Then slack should also be able to resolve the matching tensions with *certain*. Hence, invoking slack does not explain the observed difference in acceptability between the 'must $\phi \wedge$ possible $\neg\phi$' condition and the 'certain $\phi \wedge$ possible $\neg\phi$' condition. Now, as von Fintel and Gillies (2016) point out, slack can interact differently with terms that are truth-conditionally equivalent. So we could hold that, although *must* and *certain* are both maximally strong, *must* allows more pragmatic slack than *certain* (cf. von Fintel and Gillies, 2016, p. 10). To be sure, the only clear motivation for this move is to account for the patterns observed in Experiment 1, and there are reasons to think that it is not easy to independently defend. First, *certain* seems to allow for plenty of slack, as indirectly suggested by the widespread use of expressions like *absolutely certain* and *totally certain*. Second, it is generally easy to use slack to accommodate ordinary uses for all kinds of maximal adjectives such as *full* and *dry*. Third, to square the strong *must* + weak *might* package with the full pattern of results, this response would have to also be paired with the suggestion that, for auxiliaries and adjectives on the weak side of the spectrum, the tolerance for slack is reversed: i.e., possibility modals like *possible* allow more slack than *might*. As a result, it seems that, at this point, it is ad hoc to hold that the package in (5) can be defended via a construction-specific appeal to pragmatic slack.[15]

[14] Importantly, we are *not* rejecting the reasonable claim that we can expand/contract the modal base to (try to) accommodate epistemic tensions. Proponents of weak *must* + strong *might* can certainly accept this hypothesis (e.g., it might help explain why ratings in all cases are higher than expected). Still, since on this view *certain* is stronger than *must*, (altering the modal horizon to) accommodate epistemic tensions with *certain* should be harder than doing so for matching tensions with *must*. Unlike the strong view, this still predicts the difference in overall acceptability (at least under the reasonable assumption that more effortful accommodations lead to higher processing costs).

[15] As a reviewer pointed out to us, some work on imprecision with round vs. non-round numerals suggests that differences in slack susceptibility have to do with complexity. Maybe some complexity consideration can be used to explain why tensions with *certain* allow more slack than those with *must*. However, as the reviewer also points out, even if this kind of consideration ameliorates the charge the a piecemeal use of slack to defend account in (5) is ad hoc, it still requires holding that there is something about *must* that in any case makes it more susceptible to imprecision, and functionally non-maximal and weaker than *certain*; yet this move is not much consolation if the point is to save the view that *must* is maximally strong. For further discussion of this and related issues, see §4.1 below and Lassiter (2017, §6.4-6.5).

Summing up, Experiment 1A examined the degree of acceptability of epistemic tensions of various strengths. The overall pattern of results supports the hypotheses that *must* is not a maximally strong epistemic operator and that *might* is stronger than a bare possibility operator.

2.2 Experiment 1B

Experiment 1B is a variation of Experiment 1A designed to address a deflationary interpretation of part of our previous results—an interpretation which one might raise on behalf of the strong *must* + weak *might* package.

One of the key results of Experiment 1A was that tensions of the form *must* $\phi \wedge$ *possible* $\neg\phi$ were rated as significantly more acceptable than tensions of the form *certain* $\phi \wedge$ *possible* $\neg\phi$. While this result is predicted by the weak *must* + strong *might* accounts, it is in tension with accounts which stipulate that *must* is a maximally strong epistemic operator, such as the strong *must* + weak *might* account. However, proponents of strong *must* could raise the following objection. Expressions of the form 'must ϕ' might be more natural, or in some ways less awkward, than expressions of the form 'certain ϕ'. This could be simply because *must* is more frequent than *certain*; or because, independently of the effect of their interaction with the continuations/conjuncts that create epistemic tensions (in this case, *possible* $\neg\phi$), our specific vignettes can accommodate *must* more naturally than *certain*. This conjecture, when paired with the strong *must* account, could be used to explain part of the challenging pattern observed in Experiment 1A, namely, that in epistemic tensions that raise the bare possibility that $\neg\phi$, *must* ϕ was rated as significantly more acceptable than *certain* ϕ.[16]

Experiment 1B is a variation of Experiment 1A designed to test this response on behalf of the strong *must* account. As in Experiment 1A, we included tensions of the form *must* $\phi \wedge$ *possible* $\neg\phi$ and *certain* $\phi \wedge$ *possible* $\neg\phi$. However, to determine the base acceptability of *must* vs. *certain* in our vignettes, we added a 'neutral conjunct' condition. The neutral conjunct condition for each vignette was obtained by attaching a conjunct, call it *know* $\neg\psi$, to *must* ϕ or to *certain* ϕ, where ψ is selected so as to *not* raise any epistemic tension (in other words, ϕ and ψ where in each case logically and contextually independent). If the target conjecture is correct and we assume the strong *must* account, the difference in acceptability between *must* and *certain* in the epistemic tensions conditions should not be significantly greater than their difference in the neutral conjunct conditions. However, if we assume instead the weak *must* account, the difference between *must* and *certain* in the epistemic tension conditions should still be significantly greater than their difference, if any, in the neutral conjunct conditions.

[16] In principle, one could propose a similar conjecture to question the conclusion that the results of Experiment 1A support the view that *might* is stronger than other bare possibility operators (one of the two the main results of Experiment 1A which undermine the strong *must* + weak *might* package). However, in this case this conjecture is clearly less persuasive. Epistemic tensions with *might* $\neg\phi$ were *downgraded* in acceptability compared to tensions with bare possibility modals. Crucially, this was pattern was observed when comparing *might* with either of two quite different kinds of bare possibility modals (see §2.1.3), both of which are, according to standard corpus data, less frequent than *might*. For these reasons, we decided to focus on this awkwardness/frequency of usage conjectures as it applies to the *must* vs. *certain* part of the results in Experiment 1A.

For in addition to any potential baseline differences between *must* ϕ and *certain* ϕ in our vignettes, this account says that *must* ϕ, which is strictly compatible with the bare possibility that $\neg\phi$, is weaker than *certain* ϕ. As a result, a *possible* $\neg\phi$ conjunct should have an additional negative effect when conjoined with *certain* ϕ compared to its effect when conjoined with *must* ϕ.[17]

2.2.1 Materials and methods

Experiment 1B consisted of a between-subjects design where each participant rated the acceptability of an epistemic tension or neutral continuation, presented in bold, on a scale from 1 ('completely unacceptable') to 7 ('completely acceptable'). The vignette schemes used in Experiment 1B were the same as those used Experiment 1A. An example of a vignette scheme is presented in (14). We included two neutral conjunct conditions, one for each of *must* and *certain*, as illustrated in (14-a) and(14-b). We also included two conditions consisting of epistemic tensions of the form *must* $\phi \wedge$ *possible* $\neg\phi$ and *certain* $\phi \wedge$ *possible* $\neg\phi$, illustrated in (14-c) and (14-d) respectively. We should note one additional difference between the vignettes used in Experiment 1A and in this experiment. To appropriately accommodate the neutral conjunct condition, we had to eliminate a (non-rated) sentence that came after the bolded epistemic tensions in some of our original vignettes. We did this because some of these follow up sentences continued to raise epistemic tensions with respect to our target sentences (see Appendix B).

(14) I just bought a vintage bicycle at a garage sale in my neighborhood. It will need some work, but it's in decent shape. [EPISTEMIC TENSION OR NEUTRAL CONDITION HERE]

a. 'must $\phi \wedge$ not know ψ':
The bike must be from the 60s, but I don't know where it was produced.

b. 'certain $\phi \wedge$ not know ψ':
It's certain that the bike is from the 60s, but I don't know where it was produced.

c. 'must $\phi \wedge$ possible $\neg\phi$':
The bike must be from the 60s, but it's possible that it isn't.

d. 'certain $\phi \wedge$ possible $\neg\phi$':
It's certain that the bike is from the 60s, but it's possible that it isn't.

Given the 5 vignette schemes and 4 conditions for each, we had 20 critical items. Each item was evaluated by an average of 40 unique participants. We recruited 800 participants on Amazon Mechanical Turk (AMT) under the same conditions/filters

[17] In this follow-up experiment, we did not include all the conditions tested in Experiment 1A. Part of the reason for this was just that Experiment 1A was quite extensive in the amount of conditions used and number participants required to complete it. More importantly, however, our decision to compare the *must* and *certain* in the neutral conjunct conditions with 'must $\phi \wedge$ possible $\neg\phi$' and 'certain $\phi \wedge$ possible $\neg\phi$' was a deliberate conservative choice, which gave a good a priori chance to the competing hypothesis being considered in Experiment 1B. This is because the difference in acceptability between *must* ϕ and *certain* ϕ in each of the other relevant epistemic tensions—i.e., with 'possible $\neg\phi$' in the reverse order and with 'not sure that ϕ' in either order—was greater than the difference between 'must $\phi \wedge$ possible $\neg\phi$' and 'certain $\phi \wedge$ possible $\neg\phi$' in the canonical order (see §2.1.3).

used in Experiment 1A (US internet protocol addresses, approval rating greater than 90%). The experiment was deployed across one AMT assignment, which began on Wednesday, August 29th, 2019 around 8:00am PDT and terminated on Monday, September 3rd, 2019 around 9:40am PDT.

2.2.2 Predictions

The key predictions for each of our competing accounts are as follows. In this experiment, we are assuming that the strong *must* account is paired with the conjecture that, in our vignettes, *must* ϕ gets a higher base acceptability rating that *certain* ϕ. In addition, *must* is, by hypothesis, at least as strong as *certain*. Under these assumptions, the difference between *must* and *certain* in the epistemic tension conditions should be indistinguishable from their difference in the neutral conjunct conditions, which reflects their hypothesized base acceptability rates in our vignettes. This prediction is captured in (15-a), where '$|x-y| \approx |z-v|$' stands for 'the difference between x and y is indistinguishable from that between z and y'. As a result, we expect no interaction between main modal (*must* ϕ vs. *certain* ϕ) and continuation conjunct (*possible* $\neg\phi$ vs *know* $\neg\psi$), since the *possible* $\neg\phi$ continuation is not expected to have a stronger (negative) effect when conjoined with *certain* ϕ compared to when conjoined *must* ϕ.

(15) Strong *must* + weak *might*:
   a. |'must $\phi \wedge$ possible $\neg\phi$' − 'certain $\phi \wedge$ possible $\neg\phi$'|
      $\approx$ |'must $\phi \wedge$ know $\neg\psi$' − 'certain $\phi \wedge \neg$know ψ'|

The weak *must* accounts make a different prediction in this experimental set up. According to this account, *must* ϕ is weaker than *certain* ϕ, so that the former should be more suited for hosting *possible* $\neg\phi$ conjuncts. The key prediction, then, is that the difference between *must* ϕ and *certain* ϕ when conjoined with *possible* $\neg\phi$ should still be significantly greater than their (base) difference, should there be any, when each is paired with the neutral conjunct. This prediction is captured in (16-a). Accordingly, on this account, we expect an interaction between main modal (*must* ϕ vs. *certain* ϕ) and continuation conjunct (*possible* $\neg\phi$ vs *know* $\neg\psi$): specifically, *possible* $\neg\phi$ should have a stronger negative effect when conjoined with *certain* ϕ compared to on when conjoined with *must* ϕ.

(16) Weak *must* + strong *might*:
   a. |'must $\phi \wedge$ possible $\neg\phi$' − 'certain $\phi \wedge$ possible $\neg\phi$' |
      $\gg$ |'must $\phi \wedge$ know $\neg\psi$' − 'certain $\phi \wedge \neg$ know ψ'|

2.2.3 Results and discussion

The mean ratings across the 4 conditions are presented in Table 3. Using the ordinal package in R (Christensen 2015), we first conducted an ordinal mixed effects logistic regression model on the full data set, predicting acceptability ratings from a fixed effect of 'main modal' (*must* ϕ, *certain* ϕ) and a fixed effect of 'continuation conjunct' (*possible* $\neg\phi$, *know* $\neg\psi$). This model included separate, uncorrelated by-item random intercepts and slopes for main modal and continuation conjunct, respectively. We found no evidence of a significant effect of main modal

($\beta = -0.1155, SE = 0.2361, z = -0.489, p < 0.62$), but did find a significant effect of continuation conjunct ($\beta = -0.7296, SE = 0.1817, z = -4.016, p < 0.001$). Crucially, there was a significant interaction effect between main modal and continuation ($\beta = -0.5134, SE = 0.2560, z = -2.005, p < 0.05$). We next conducted pairwise ordinal mixed effects logistic regression models predicting acceptability ratings between the epistemic tensions conditions, 'must $\phi \wedge$ possible $\neg\phi$' condition and the 'certain $\phi \wedge$ possible $\neg\phi$, and also between the neutral conjunct conditions, 'must $\phi \wedge \neg$know ψ' condition and the 'certain $\phi \wedge \neg$know ψ'. This model included by-item random intercepts and by-item random slopes for condition. We found a significant difference between the 'must $\phi \wedge$ possible $\neg\phi$' condition and the 'certain $\phi \wedge$ possible $\neg\phi$' condition ($\beta = -0.561, SE = 0.193, z = -2.901, p < 0.004$), thereby replicating one of the main findings of Experiment 1A. Importantly, we found no significant difference between the neutral conjunct conditions, i.e., between 'must $\phi \wedge \neg$know ψ' condition and the 'certain $\phi \wedge \neg$know ψ' ($\beta = -0.1174, SE = 0.269, z = -0.437, > 0.66$).

Condition	**Mean Rating**
must $\phi \wedge$ not know ψ	5.50 (SD = 1.71)
certain $\phi \wedge$ not know ψ	5.53 (SD = 1.43)
must $\phi \wedge$ possible $\neg\phi$	4.84 (SD = 1.72)
certain $\phi \wedge$ possible $\neg\phi$	4.23 (SD = 1.95)

Table 3 Results of Experiment 1B. Although there was no difference between 'must $\phi \wedge$ not know ψ' and 'certain $\phi \wedge$ not know ψ' (i.e, between the 'neutral conjunct' conditions), epistemic tensions of the form 'must $\phi \wedge$ possible $\neg\phi$' obtained significantly better ratings than those of the form 'certain $\phi \wedge$ possible $\neg\phi$'.

The basic pattern of results observed in Experiment 1B is straightforward, given the aim of this experiment. Although we replicated one of the main results of Experiment 1A, namely, a significant difference such that epistemic tensions of the form 'must $\phi \wedge$ possible $\neg\phi$' were rated as significantly better than tensions of the form 'certain $\phi \wedge$ possible $\neg\phi$', there was no significant difference between the neutral continuation conditions, i.e., between 'must $\phi \wedge$ not know ψ' and 'certain $\phi \wedge$ not know ψ'. This distinction confirms prediction (16-a) of the weak *must* account, and disconfirms prediction (15-a) of the strong *must* account. This difference drove the interaction effect uniquely predicted by the weak *must* account: namely, the interaction between the main modal (*must* ϕ vs. *certain* ϕ) and the continuation conjunct (*possible* $\neg\phi$ vs. $\neg$ *know* ψ). Accordingly, these results not only replicate one of the main findings of Experiment 1A, but also strongly suggest that, in our vignettes, *certain* is not, in itself, significantly disfavored relative to *must*. The best explanation of the overall pattern, given our competing accounts, is that *must* is a non-maximal epistemic operator, one that is in particular weaker than *certain*, and is thus more tolerant to being placed in contexts of epistemic tension.

Summing up, Experiments 1A-B provide substantial evidence—based on various contexts of epistemic tension and the corresponding comparative acceptability ratings—against the strong *must* + weak *might* package, and in favor of the weak *must* + strong *might* packages. Still, as we mentioned before, Experiments 1A-B can't be used to distinguish between our competing accounts of weak *must*, for

both entail that *must* ϕ is (i) strictly compatible with the possibility that not ϕ, and (ii) weaker than matching assertions with *certain* ϕ.

3 Experiments 2A-B

In §1-1.2 we introduced two accounts that incorporate a weak semantics for *must*. One is a probabilistic account according to which *must* ϕ expresses high but non-maximal likelihood that ϕ. The other is a Kratzer-style restricted quantificational account according to which *must* ϕ expresses that ϕ holds in all the best worlds of the epistemic modal base. We also pointed out that only the probabilistic implementation is compatible with the pure epistemic language hypothesis. These accounts can't be distinguished on the basis of Experiments 1A-B. Experiments 2A-B, to which we now turn, focus on cases where these accounts do make different predictions.

The target cases are similar to (2), i.e., scenarios where downplaying a previous *must*-claim whose prejacent turns out to be false feels odd or unjustified. These cases were first introduced by von Fintel and Gillies (2010, 2016), who presented them as objections against weak (non-veridical) *must* in general. The objection turns on the intuitive acceptability of downplaying a previous high (but non-maximal) certainty claim vs. the oddness of downplaying a previous *must*-claim in otherwise matching scenarios. Consider the dialogues in (17) and (18). According to von Fintel & Gillies, (17-c), which downplays a previous high (but non-maximal) certainty claim, is intuitively fine. In contrast, (18-c), which downplays a previous *must*-claim, feels substantially worse.

(17) a. Alex: Its 98% certain that its raining.
b. Billy: [*opens curtains*] Not it isn't. You were wrong.
c. Alex: Well, strictly speaking, I was not wrong. I was careful. I only said that it was 98% certain that it is raining.

(18) a. Alex: It must be raining.
b. Billy: [*opens curtains*] Not it isn't. You were wrong.
c. Alex: ?? Well, strictly speaking, I was not wrong. I was careful. I only said that it must be raining.

Suppose the perceived difference between (17-c) and (18-c) represents a general contrast. What is the theoretical import of this result? In particular, does it undermine accounts of weak (non-veridical) *must* in general, as von Fintel and Gillies (2010, 2016) seem to hold? Following standard accounts of the norms of assertion, let us assume that speakers are generally committed to knowing the content determined by the literal meaning of their assertions (Williamson, 2000; Lauer, 2013).[18] Accordingly, we expect that intuitions about the degree of acceptability or justifiability of downplaying claims should reflect the following principle: the stronger the original modalized claim, the harder, or less justified, it will be to

[18] As we discuss in detail in Experiment 2B (see below), the previous assumption about speakers' commitments to the content of their assertions, at least in the version we endorse, does not to exclude the possibility that speakers are also sometimes committed to the implicatures of their assertions, at least when the latter are not explicitly cancelled (see Klecha, 2018). We briefly discuss other norms of assertion in §4.1.

subsequently downplay it if the prejacent turns out to be false. From this perspective, it is easy to see that the contrast between (17-c) and (18-c) would challenge the probabilistic account of weak *must*, as acknowledged by both defenders and critics of this account. For according to the probabilistic account, *must* ϕ entails that the likelihood of ϕ is above some high but non-maximal threshold. It follows that downplaying a previous *must* ϕ claim should be roughly as acceptable as downplaying a matching overt claim of high but non-maximal certainty in ϕ. Yet this prediction is in tension with the perceived contrast between (17-c) and (18-c): downplaying a claim of high (but non-maximal) certainty feels substantially more justified, or at least less odd, than downplaying a *must*-claim.[19]

Crucially, however, the difference between (17-c) and (18-c) is expected under the restricted quantificational account of weak *must*. To see why, consider the role of ordering sources and normality assumptions. The ordering source for *must* and *might*, recall, is determined by a stereotypical conversational background, g_s, which picks out sets of relevant assumptions which capture what is taken to be normal at each world. In the contexts where dialogues like (17) and (18) take place, normality assumptions include information like 'that the laws of physics won't suddenly change' and, supposing Alex used evidence from a weather channel report, 'that if the weather channel says it is rain at l, then it is raining at l'. That is, these are assumptions that interlocutors normally believe, or take for granted, in particular contexts of deliberation. To capture this feature of normality assumptions, we will adapt a related proposal due to Swanson (2016), and assume that ordering sources for *must* and *might* have to satisfy a specific doxastic constraint, as spelled out in (19):

(19) **Stereotypical conversational background.** A function g_s such that for any world w, $g_s(w)$ represents what is normal in w according to some *suitable standard* in w.

a. A set of premises P *suitably represents* what is normal in w only if, for each $p \in P$, $B_S(p)$, where S includes at least the speaker (or relevant subjects in some embedded cases).

Given (19), it follows that, when asserting a *must*-claim, speakers commit themselves to *believing*, for the purposes at hand, the normality assumptions.[20] Crucially, this does not require that speakers commit themselves to *knowing* these assumptions: as Kratzer emphasizes, normality assumptions are after all defeasible.[21] Still, assuming the restricted quantificational account and (19), we can show that *must*-claims have full doxastic strength with respect to their bare prejacent. For by asserting *must* ϕ, a speaker S is now committed to (i) knowing the relevant

[19] Some readers might worry that, even if they are roughly equivalent in terms of their logical force, *almost certain* and *must* could generate different implicatures, which might in turn differentially affect whether they can tolerate downplaying moves. Since Experiment 2B, presented in §3.2 below, was designed in par to explore this possibility, we will discuss these issues in detail that section.

[20] In §4.3, we discuss some proposals for relaxing or even dropping this kind of doxastic constraint on the stereotypical ordering sources used by *must* and *might*, while preserving the virtues, with respect to downplaying cases, of the restricted quantificational account.

[21] That *must* semantically encodes a doxastic but not an epistemic constraint on normality assumptions is part of what distinguishes this version of the restricted quantificational account from the view that *must* is a maximally strong epistemic operator.

evidence, (ii) believing the relevant set of normality assumptions, and (iii) knowing that ϕ follows from (i) and (ii). Note that premise (ii) follows from our doxastic constraint on normality assumptions, and premise (iii) from the semantics of *must*, under the restricted quantificational account in (7). From (i)-(iii) we can derive the result that an assertion by S of *must* ϕ entails $B_S(\phi)$. In contrast, asserting that ϕ is 'almost certain' only commits S to believing that ϕ has a high likelihood, given the evidence, but doesn't entail any full or unhedged doxastic commitment to the bare prejacent. Due to this difference in their doxastic entailments, it follows that *must* ϕ should be harder to downplay than 'almost certain that ϕ'.[22]

At this point, we can conclude that epistemic tensions in downplaying contexts such as (17) and (18) have the potential to distinguish between the probabilistic and restricted universal accounts of weak *must* (the full predictions of each account are presented in detail in §3.1.2). Still, there are three reasons why the acceptability judgments should, given the state of the debate, be obtained in controlled experimental settings. Firstly, the target judgments are subtle. For example, Lassiter (2016) rejects the intuitions of von Fintel and Gillies (2010, 2016), and submits that downplaying a previous claim of high but non-maximal certainty, as in (17), feels as odd/unjustified as downplaying a previous *must*-claim, as in (18). If this is correct, downplaying cases do not undermine—and actually support—the weak probabilistic account of *must*. Secondly, the debate has focused on cases essentially like (17)-(18), so we don't know whether the perceived contrast, if there is one, generalizes. Thirdly, we can add, as experimental conditions, versions of downplaying scenarios that allow us to test whether the strong account really makes the right predictions. In short, although downplaying scenarios have the potential to inform these debates, the target acceptability patterns should at this point rest on systematic empirical investigation. Experiments 2A-B are a first step in that direction.

[22] The view that the normality assumptions for *must* have a doxastic but not an epistemic constraint presupposes a doxastic and epistemic logic that allows agents to (coherently) reflectively believe propositions that they do not believe they know. This is not just a matter of belief being non-veridical; it is also about the compatibility of believing that ϕ while acknowledging that it is strictly possible that not ϕ. Technically, this is a trivial requirement, since various recent logics for the K_i and B_i-operators already respect these properties. For example, van Benthem (2010) and van Benthem and Smets (2015) construe B_i as a universal quantifier over the 'most plausible' worlds of epistemic spaces (were we assume that if g_p picks out the premises for a plausibility ordering, and g_s for a stereotypical ordering source, then for each $w \in W$, $g_s(w) \subseteq g_p(w)$; that is, plausibility orderings can include more information). From this perspective, even if an assertion by S of *must* ϕ entails that $B_S(\phi)$, this is consistent with S acknowledging that it is strictly possible that $\neg\phi$. Importantly, this ensures that our account of the doxastic strength of *must*-claims is not in tension with the results of Experiment 1, according to which *must* ϕ can be conjoined with various expressions of the bare possibility that $\neg\phi$. The role of plausibility orderings in the logic for belief also clarifies precisely why, on the current view, an assertion by S that ϕ is 'almost certain' doesn't entail $B_S(\phi)$. For there are situations where ϕ can have a very high likelihood, given the evidence, and yet ϕ does not hold in all of the most plausible worlds. Typical examples of this kind involve fair lottery cases (often used to argue against simple 'Lockean' theories of belief, see e.g., Harman 1986). Suppose John holds one ticket for a fair lottery, knows that he has a very low chance of winning, and doesn't yet know the results of the draw. John can then assert that he is 'almost certain' he'll loose, and believe that it is very very likely that he'll loose, without also believing that he lost.

3.1 Experiment 2A

3.1.1 Materials and methods

Experiment 2A consisted of a between-subjects design where each participant rated the acceptability of a 'downplaying' response presented within a short dialogue. Each downplaying response was presented in bold, and participants were asked to rate it on a scale from 1 ('completely unjustified') to 7 ('completely justified'). Across five distinct dialogue contexts, we tested the perceived acceptability of downplaying a previous epistemically modified assertion in three conditions: a 'must' condition, an 'almost certain' condition, and a 'certain' condition. A representative dialogue scheme is presented in (20) (background information is italicized), and the three target conditions are schematically presented in (21-a)-(21-c):

(20) *Alex and Billy are deliberating about whether to go outside. They will do so only if it is* not *raining. Alex looks at two weather forecasting websites, and both say it is raining at their location.*

Alex: [MODAL + it's raining outside].
Billy opens the curtains. Billy and Alex can both see it is actually not raining outside.
Billy: No it isn't. You were wrong.
Alex: **Well, strictly speaking, I was not wrong. I only said that [MODAL + it's raining outside].** Anyway, lets go out then.

(21) a. 'Must' condition:
Alex: It must be raining outside. ... Alex: **Well, strictly speaking, I was not wrong. I only said that it must be raining outside.** ...

b. 'Almost certain' condition:
Alex: It is 95% certain that it's raining outside. ... Alex: **Well, strictly speaking, I was not wrong. I only said that it's 95% certain that it's raining outside.** ...

c. 'Certain' condition:
Alex: It is certain that it's raining outside. ... Alex: **Well, strictly speaking, I was not wrong. I only said that it's certain that it's raining outside.** ...

As illustrated in (21-a)-(21-c), the 3 conditions in Experiment 2 involved downplaying a prior *must* claim, a prior claim expressing very high but non-maximal certainty, and a prior claim of unhedged/full certainty. We tested each of these 3 conditions across 5 dialogue contexts, resulting in a total of 15 critical stimuli (see Appendix C), for each of which we obtained 40 responses.

We recruited 600 participants on Amazon Mechanical Turk (AMT). As with the previous experiments, we recruited only AMT workers with US internet protocol addresses who possessed an approval rating greater than 90% for past tasks completed on AMT. The experiment was deployed across several AMT assignments (HITs), which were all simultaneously published on September 04, 2017 around 1:00 pm PDT and which had concluded by September 11, 2017 around 4:00 pm PDT. To ensure that workers did not complete our experiment more than once,

we used the UniqueTurker script (https://uniqueturker.myleott.com/) as well as the warning message used on the other experiments. One worker attempted to complete four HITs, all of which were excluded from the analysis. Users who had participated in prior versions of a norming experiment, or who had participated in pilot versions of the present experiment, were blocked from participating.

3.1.2 Predictions

To begin, note that all the accounts predict that downplaying in the 'almost certain' condition should be more acceptable than in the 'certain' condition, as captured in (22), where '$x \succ y$' means that 'x is more justified than y'. The reason is that, in general, the weaker an assertion the easier it should be to subsequently downplay it. This might seem like a trivial prediction, but confirming it provides evidence that our experimental task was executed by participants as intended. In addition, Experiment 2 includes a 'certain' condition. This allows us to again test the strong *must* account. Although the results of Experiment 1 undermined this account, Experiment 2 could further corroborate those results using a different paradigm.

(22) General prediction:
  a. almost certain condition $\succ$ certain condition

Let us now consider the unique predictions of each account, starting with those of the strong *must* account, summarized in (23). On this account, *must* is maximally strong, so downplaying a previous *must*-claim should be hard. Specifically, downplaying a *must*-claim should be less acceptable than downplaying a high but non-maximal certainty claim. Hence the 'must' condition is predicted to be less acceptable than the 'almost certain' condition, as captured in (23-a). In addition, on this strong view *must* ϕ entails that ϕ is certain. Now, that *must* is a maximally strong epistemic operator is on this account non-negotiable, whereas strong theorists could hold that *certain* is equal or less strong than *must*. As a result, downplaying in the 'must' condition is predicted to be either equally or less acceptable than downplaying in the 'certain' condition, as captured in (23-b).

(23) Strong *must*:
  a. almost certain condition $\succ$ must condition
  b. certain condition $\succ$ / $\approx$ must condition

Consider next the unique predictions of the probabilistic account of weak *must*, summarized in (24). On this view, *must* ϕ entails that $Pr(\phi) > \theta_c$, where θ_c is a high but non-maximal probability. As a result, downplaying in the 'must' condition is predicted to be more acceptable than in the 'certain' condition, as captured in (24-a). Furthermore, as pointed out by von Fintel and Gillies (2010, 2016) and Lassiter (2016), these accounts predict that the 'must' condition and the 'almost certain' condition should be roughly equally tolerant to downplaying. Indeed, simple probabilistic accounts arguably predict that the 'must' condition should be more tolerant or acceptable. There are two reasons for this. First, the threshold θ_c is context-sensitive. Secondly, given the kinds of contexts that license *must*-claims—see e.g. the corpus data presented in Lassiter (2016, 2017)—we can't hold that, in general, θ_c is extremely high and inflexible. Indeed, it is arguable that,

on this view, participants should fix θ_c by applying charity and hence increase the acceptability of downplaying responses. This prediction is captured in (24-b).

(24) Weak probabilistic *must*:
  a. must condition $\succ$ certain condition
  b. must condition $\succ$ / $\approx$ almost certain condition

Finally, consider the predictions of the restricted universal account of weak *must*, summarized in (25). According to this account, *must* ϕ entails that ϕ holds in the best worlds of the epistemic modal base. In everyday joint deliberation scenarios like (20), the ordering source used to determine the best worlds will include reasonable assumptions defeasibly believed by speakers. As a result, *must* is weaker than *certain*, so downplaying in the 'must' condition is predicted to be easier than in the 'certain' condition, as captured in (25-a). On the other hand, *must* ϕ commits speakers to believing that ϕ, i.e., to taking it as an unhedged premise in deliberation/decision. In contrast, asserting that ϕ has a high but non-maximal certainty does not as such commit the speaker to believing that ϕ, i.e., to holding that, for the relevant practical purposes, bare ϕ can be assumed as an unhedged premise in subsequent deliberation. As a result, downplaying in the 'almost certain' condition should be more acceptable than in the 'must' condition, as captured in (25-b).

(25) Weak universal *must*:
  a. must condition $\succ$ certain condition
  b. almost certain condition $\succ$ must condition

3.1.3 Results and discussion

The average ratings across the three downplaying conditions are presented in Table 4. Ratings were based on a 1 ('completely unjustified') to 7 ('completely justified') scale. We used mixed-effects ordinal logistic regression models to analyze the 'rate the justifiability' responses with by-vignette random intercepts and by-vignette random slopes for condition. For our initial analysis, the '95% certain' discourses (2 vignettes) and 'almost certain' discourses (3 vignettes) were coded together as the 'almost certain' condition, and the 'must' condition was used as a reference level. The difference between the 'must' and 'almost certain' conditions was significant ($\beta = 1.140$, $SE = 0.217$, $z = 5.251$, $p < 0.0001$), as was the difference between the 'must' and 'certain' conditions ($\beta = -0.794$, $SE = 0.278$, $z = -2.857$, $p < 0.005$). We next performed pairwise logistic regression analyses to confirm that the '95% certain' discourses and the 'almost certain' discourses each individually outperformed the corresponding 'must' condition. There was a significant differences between the '95% certain' condition and the 'must' condition in matching vignettes ($\beta = -1.194, SE = 0.293, z = -4.08, p < 0.0001$). There was also a significant difference between the 'almost certain' condition and the 'must' condition in matching vignettes ($\beta = -1.229, SE = 0.239, z = -5.142, p < 0.0001$). Note that because only a subset of the five vignettes were analyzed in each of these models, we did not include by-vignette random effects.

The basic pattern of results is straightforward. To begin, note that the observed pattern confirms the prediction—shared by all accounts and which we take

Condition	Mean Rating
almost certain ϕ	4.63 (SD = 1.77)
must ϕ	3.39 (SD = 1.83)
certain ϕ	2.73 (SD = 1.88)

Table 4 Average ratings per condition for Experiment 2A. Downplaying in the 'almost certain' condition was significantly more acceptable than in the 'must' condition. In addition, downplaying in the 'must' condition was significantly more acceptable than in the 'certain condition'.

as evidence that the task was performed by participants as intended—that the 'almost certain' condition should be rated as significantly more acceptable than the 'certain' condition. Next, note that the pattern directly corresponds to the predictions of the restricted universal account of weak *must*, outlined in (25). Specifically, downplaying in the 'must' condition was rated as more acceptable than in the 'certain' condition, confirming prediction (25-a), and downplaying in the 'almost certain' condition was rated as significantly more acceptable than in the 'must' condition, confirming prediction (25-b).

In addition, the results disconfirm unique predictions of both the strong and the weak probabilistic accounts of *must*. Let us begin with the latter.

The results disconfirm the unique prediction (24-b) of the probabilistic account of weak *must*. Downplaying in the 'must condition' was rated as significantly less acceptable than in the 'almost certain condition'. Lassiter (2016) and von Fintel and Gillies (2010, 2016) agree that prediction (24-b) distinguishes weak probabilistic accounts. However, while Lassiter expects that (24-b) will be observed, von Fintel & Gillies predict that it won't. Experiment 2A confirms von Fintel & Gillies' prediction. Indeed, the problem for probabilistic accounts of weak *must* is considerable. As we mentioned above, (i) weak probabilistic accounts are committed to the view that the relevant likelihood threshold is context-sensitive. At the same time, (ii) corpus data such as that used (to construct stimuli) in Experiments 1A-B suggest that, in normal everyday contexts, *must* claims are acceptable even if there is substantial uncertainty. In addition, (iii) when fixing context-sensitive parameters participants arguably tend to use charity. Taken together, (i)-(iii) suggest that the simple probabilistic account predicts that the 'must' condition should be more tolerant of downplaying than the 'almost certain' condition. The results of Experiment 2A, however, display precisely the inverse relation between these conditions.[23]

[23] Yalcin (2011) and Knobe and Yalcin (2014) argue that judgments about the adequacy of retractions of (prior) epistemically hedged claims in cases where the bare prejacent turns out to be false are sensitive, perhaps primarily, to whether (after the factual revelation) updating the common ground or relevant information state with the original epistemically hedged assertion is still appropriate (see also Khoo 2014). It is likely that this kind of factor does have (some) negative effect on ratings in our downplaying scenarios: specifically, that participants punish speakers for choosing to downplay instead of retract the original modally hedged claims. The question of interest, for us, is whether this perspective can help proponents of the weak probabilistic account explain our results (cf. Lassiter, 2016, p.143-4). This doesn't seem promising. In all of our conditions, the potential updates are inadvisable (or pointless) after the falsity of the prejacent is revealed (at least if, for simplicity, we ignore implicit tense). Still, the ratings are clearly sensitive to the modal strength of the original assertion, as can be seen by comparing the 'almost certain' condition with the 'certain' condition. Furthermore, given the semantics postulated by the weak probabilistic account, updates with 'almost certain that ϕ' and 'must ϕ' are, in general, expected to be roughly equally strong or informative; however,

A surprising result of Experiment 2A—given the original dialectical use of downplaying tensions (e.g., in von Fintel and Gillies, 2010, 2016)—is the disconfirmation of prediction (23-b) of the strong *must* account. The prediction is that, since *must* is maximally strong, downplaying a previous *must* ϕ assertion should be at least as unjustified as downplaying a matching assertion that ϕ is certain (for recall that strong *must* theorists can relax the assumption that *certain* is maximally strong, but obviously not that *must* is). The results show, however, that downplaying in the 'must' condition was significantly more acceptable than in the 'certain' condition. This undermines the view that *must* is maximally strong, corroborating the results of Experiments 1A-B. In short, the downplaying paradigm first used by von Fintel and Gillies (2010) to argue against weak *must* doesn't directly support their account. von Fintel and Gillies focus on the perceived difference between downplaying a *must*-claim vs. a claim of high but non-maximal certainty. This contrast *was* confirmed by our results. At the same time, they don't consider whether there is also a contrast between downplaying a *must* vs. a *certainty* claim. On their view, the former should be at least as bad as the latter. However, the results of Experiment 2A disconfirm this prediction: downplaying in the 'must' condition was significantly more acceptable than in the 'certain' condition.

3.2 Experiment 2B

Experiment 2B is a variation of 2A designed with two goals in mind: first, to determine whether the main results of Experiment 2A replicate over some variation in the target stimuli; second, to control for two potential confounds of Experiment 2A.

In Experiment 2A we asked participants to rate whether downplaying statements of the form 'I only said that MODAL + ϕ' felt justified. *Only* is a focus sensitive operator which arguably requires that there be at least one formal alternative not entailed by its prejacent (Rooth, 1992, 1996). In light of this, consider again the 'must' condition. One could argue that the only formal alternative to *must* ϕ is *might* ϕ. So even if *must* is probabilistic and non-maximal, configurations of *only* + *must* ϕ could be downgraded in acceptability because there is no formal alternative that can be negated. In contrast, the stimuli used in the 'almost certain' condition where of the form 'I only said that it is almost certain that ϕ' and 'I only said that it is 95% certain that ϕ', and both clearly have stronger and salient formal alternatives. From this perspective, the observed downgrading in the ratings for the downplaying responses in the 'must' condition relative to the 'almost certain' condition has a reasonable explanation which is independent of the specific semantics for *must*. To control for this, the stimuli used in Experiment 2B mention the bare prejacent ϕ just before the downplaying claim 'I only said that MODAL + ϕ', as illustrated in (26) below. In this way, we turn the prejacent ϕ into an active, contextually relevant alternative, even if (on some accounts) it is not strictly a formal alternative to MODAL + ϕ.[24]

what we have to explain is why 'almost certain' claims are substantially easier to downplay than *must*-claims.

[24] One could defend our original design by arguing that since ϕ is structurally simpler and arguably formally derivable from *must* ϕ, there is reason to think that it is in general an active

In Experiment 2A our stimuli for the 'almost certain' condition were of two kinds: 'it is almost certain that ϕ' and 'it is 95% certain that ϕ'. As the previous discussion suggests, these sentences could give rise to upper bounded scalar implicatures, resulting in strengthened readings roughly like 'it is almost certain that $\phi \wedge \neg$ it is certain that ϕ' and 'it is 95% certain that $\phi \wedge \neg$ it is more than 95% certain that ϕ'. Furthermore, assume that participants don't derive analogous upper bounded scalar implicatures in the 'must' condition (because the bare prejacent of a *must*-claim is not a (salient) alternative). This would explain why (enriched) assertions in the 'almost certain' condition are interpreted as weaker, hence easier to subsequently downplay, than assertions of *must* ϕ. This holds even if we assume that *must* is weak and probabilistic. To control for this, Experiment 2B includes two versions of the 'almost certain' condition. One is the original '95% certain that ϕ' condition. In this way, we can see whether the main results of Experiment 2A replicate. The other condition includes stimuli of the form 'at least 95% certain that ϕ'. By introducing the 'at least' modifier, we block the computation of upper bounded implicatures (Krifka, 1999; Mayr, 2013), and thus control for this potential confound. Should the original contrast between the 'must' and 'almost certain' conditions still hold, proponents of the weak probabilisitic account can't plausibly argue that this is due to the presence of upper bounded implicatures in the 'almost certain' condition.

3.2.1 Materials and Methods

Experiment 2B also used a between-subjects design. The task and rating scale used for Experiment 2B were the same as for Experiment 2A. A representative dialogue scheme used in Experiment 2B is presented in (26), and the target conditions are presented in (27-a)-(27-d).

(26) *Alex and Billy are deliberating about whether to go outside. They will do so only if it is* not *raining. Alex looks at two weather forecasting websites, and both say it is raining at their location.*

Alex: [MODAL + it's raining outside].
Billy opens the curtains. Billy and Alex can both see it is actually not raining outside.
Billy: No it isn't. You were wrong.
Alex: **Well, strictly speaking, I was not wrong. I didn't say that it's raining. I only said that [MODAL + it's raining outside].** Anyway, let's go out then.

(27) a. 'Must' condition:
Alex: It must be raining outside. ... Alex: **Well, strictly speaking, I was not wrong. I didn't say that it's raining. I only said that it must be raining outside.** ...

b. 'Almost certain' condition (as before):
Alex: It is 95% certain that it's raining outside. ... Alex: **Well,**

alternative to assertions of *must* ϕ (cf. Mandelkern, 2016; Goodhue, 2017). We are sympathetic to this response, but concluded that it is nevertheless valuable to determine whether the results of Experiment 2A would replicate using a version of the downplaying cases in which the bare prejacent is explicitly introduced into the relevant context.

strictly speaking, I was not wrong. I didn't say that it's raining. I only said that it's 95% certain that it's raining outside. ...

c. 'Almost certain' condition (SI blocked):
Alex: It is at least 95% certain that it's raining outside. ... Alex: **Well, strictly speaking, I was not wrong. I didn't say that it's raining. I only said that it's at least 95% certain that it's raining outside.** ...

d. 'Certain' condition:
Alex: It is certain that it's raining outside. ... Alex: **Well, strictly speaking, I was not wrong. I didn't say that it's raining. I only said that it's certain that it's raining outside.** ...

As illustrated in (27-a)-(27-d), the 4 main conditions in Experiment 2 involved: downplaying a prior *must*-claim, downplaying a prior claim expressing very high but non-maximal certainty, in two versions, one as in Experiment 2A and one with the addition of an 'at least' modifier, and downplaying a prior claim of unhedged or full certainty. We tested each of these conditions across the 5 dialogue contexts, resulting in a total of 20 critical stimuli (see Appendix D), for each of which we obtained 40 responses.

We recruited 800 participants on Amazon Mechanical Turk (AMT). We recruited only AMT workers with US internet protocol addresses who possessed an approval rating greater than 90% for past tasks completed on AMT and who accessed AMT via a United States IP Address. The experiment was deployed across several AMT assignments (HITs), which were all simultaneously published on January 29, 2018 around 10:00 am PDT and which had concluded by February 04, 2018 around 11:00 am PDT. To ensure that workers did not complete our experiment more than once, we used the UniqueTurker script (https://uniqueturker.myleott.com/) as well as the warning message used on the other experiments. These two safeguards were successful: no worker participated in more than one HIT. We allowed workers to participate in this experiment if they had participated in a prior version of our other experiments which were run roughly five months prior.

3.2.2 Predictions

The predictions of the competing accounts are the same as in Experiment 2A: the predictions of the strong *must* are as in (23), the predictions of the probabilistic account of weak *must* are as in (24), and the predictions of the weak universal account of *must* are as in (25). The key thing to note is that Experiment 2B has two versions of the 'almost certain' condition: one of the form '95% certain that ϕ' and one of the form 'at least 95% certain that ϕ'.

3.2.3 Results and discussion

The average rating across the three downplaying conditions are presented in Table 5. As in Experiment 2A, we used a mixed-effects ordinal logistic regression model to analyze the 'rate the justifiability' responses with by-vignette random intercepts and by-vignette random slopes for condition and 'must' as the reference level. The difference between the '95% certain' and the 'must' conditions was

significant ($\beta = 1.06, SE = 0.218, z = 4.84, p < 0.0001$), as was the difference between the 'must' and the 'certain' conditions ($\beta = -0.976, SE = 0.325, z = -3.01, p < 0.003$). A subsequent pairwise analysis revealed that the difference between the '95% certain' and 'at least 95% certain' conditions was not significant ($\beta = -0.100, SE = 0.193, z = -0.520, p < 0.61$).

Condition	**Mean Rating**
at least 95% certain ϕ	4.65 (SD = 1.94)
95% certain ϕ	4.76 (SD = 1.89)
must ϕ	3.63 (SD = 1.97)
certain ϕ	2.68 (SD = 1.74)

Table 5 Average ratings per condition for Experiment 2B. Downplaying in the 'at least 95% certain' condition was significantly more acceptable than in the 'must' condition. And downplaying in the 'must' condition was significantly more acceptable than in the 'certain' condition.

The basic pattern of results observed in Experiment 2B is straightforward. Downplaying in either version of the 'almost certain' condition was rated as significantly more acceptable than in the 'must' condition. This disconfirms prediction (24-b) of the weak probabilistic account, and confirms prediction (25-a) of the restricted universal account. In addition, downplaying in the 'must condition' was in turn rated as significantly more acceptable than in the 'certain condition'. This disconfirms prediction (23-b) of the strong account, and confirms prediction (25-b) of the restricted universal account. Overall, the observed pattern of results in Experiment 2B replicates the results obtained in Experiment 2A. As before, this pattern is only predicted by the restricted quantificational account of *must*.

In addition to replicating our original results, Experiment 2B controls for two potential confounds of Experiment 2A. The first is that, since in the original stimuli we didn't mention the bare prejacent in the discourse context, assertions of 'I only said that must ϕ' could be downgraded because of the unavailability of a stronger salient alternative. The second is that assertions of the form 'it is almost certain that ϕ' and 'it is 95% certain that ϕ' could give rise to upper-bounded implicatures. Each of these potential confounds could be invoked on behalf of the probabilistic weak *must* account to square prediction (24-b) with the original result that downplaying previous assertions of high but non-maximal certainty in ϕ was rated as substantially more justified than previous assertions that *must* ϕ. To control for this, the stimuli used in Experiment 2B included two modifications: (i) the bare prejacent was mentioned in the context immediately preceding the downplaying move, and (ii) we included a condition where 'at least' modifies the claim of high but non-maximal certainty, thereby blocking potential upper bounded implicatures. Given these controls, and the observed replication of our original results, we conclude that the evidence against prediction (24-b) of the probabilistic account of weak *must* is robust.

To sum up, Experiment 2A-B examined whether participants felt that it was justified to downplay a previous assertion of the form 'modal + ϕ', in contexts where ϕ turned out to be false. The target epistemic modals varied in strength from high but non-maximal to unambiguously maximal. The pattern of results, we have argued, undermines two reasonable hypotheses about the logical strength

of epistemic *must*, both compatible with the pure epistemic language hypothesis. The results suggest that *must* is neither a maximally strong epistemic operator (cf. von Fintel and Gillies, 2010), nor a high but non-maximal probabilistic one (cf. Lassiter, 2016). They support the view that *must* involves universal quantification over the best worlds of epistemic modals bases (cf. Kratzer, 2012; Roberts, 2015; Giannakidou and Mari, 2016). Corroborating the results of Experiments 1A-B, Experiment 2A-B also suggests that *certain* is a stronger epistemic operator than *must*.

4 General Discussion

Experiments 1A-B and 2A-B pitted the mantra on *must* and *might* against two recent accounts which are compatible with the pure epistemic language hypothesis (cf. von Fintel and Gillies, 2010, 2016; Lassiter, 2016, 2017). Our experiments were designed to obtain systematic and comparative acceptability ratings for various kinds of epistemic tensions. We tested tensions previously discussed in this debate, as well as novel variations that, we argued, are crucial to make further progress. The overall pattern of results clearly favors the mantra over at least the simple or unmodified versions of its recent competitors (Kratzer, 1981, 2012; Roberts, 2015; Giannakidou and Mari, 2016).

The main shortcoming of the strong *must* + weak *might* package, brought out in Experiments 1A-B, is that it incorrectly predicts that, in contexts that admit the strict possibility that $\neg\phi$, assertions of *must* ϕ should be as unacceptable as assertions of maximal certainty in ϕ. It also incorrectly predicts that assertions that *might* ϕ should be as weak as assertions of the bare possibility that ϕ. The main shortcoming of the probabilistic weak *must* + strong *might* account (cf. Lassiter, 2016), brought out in Experiments 2A-B, is that it incorrectly predicts that, when a previous assertion of *must* ϕ turns out to be wrong, we should be as justified in downplaying it as we are in downplaying matching assertions of very high but non-maximal certainty that ϕ.

Thus far, we have focused on simple versions of each competing account. We also maintained the initial assumptions on the semantics of *certain* and *possible*, including that their modal base is determined in context in the same way as for the epistemic auxiliaries. Are there reasonable departures from these simple versions of the competing accounts, or from any of our fixed background assumptions, that help reconcile these accounts with our results? The aim of this General Discussion is to explore some revisions and refinements of the competing accounts. This will, in turn, help us highlight which components of the mantra are responsible for its success vis-á-vis our contexts of epistemic tension, and suggest some promising avenues for future research.

4.1 Revisiting the strong *must* + weak *might* package

If we assume that *must* is a maximally strong epistemic operator, and that *must* ϕ entails ϕ (von Fintel and Gillies, 2010, 2016), it is hard to make sense of the result, supported by Experiments 1A-B, that assertions of *must* ϕ are strictly compatible with the possibility that $\neg\phi$. As Lassiter (2016) points out, one way of trying to

reconcile veridical *must* with this observation is to adopt a weak norm of assertion. For example, instead of the knowledge norm (K-A) (Williamson, 2000), we could adopt the proper justification (belief) norm (J-A) (cf. Lackey, 2007; Kvanvig, 2009):

(K-A) Assert p only if you know that p.

(J-A) Assert p only if you have proper justification for (believing that) p.

Given (J-A), asserting that *must* ϕ need not commit the speaker to the truth of ϕ—even under a veridical account of *must*—but only to having proper justification for ϕ. This latter commitment is compatible with the speaker's also believing that it is strictly possible that $\neg\phi$. Thus, combining the strong *must* + weak *might* package with a norm of assertion such as (J-A) seems to allow for the acceptability of epistemic tensions such as (1).[25]

Does this response help reconcile veridical *must* with the *full* pattern of results obtained in Experiments 1A-B? Consider the comparative acceptability ratings. What needs to be explained is why each epistemic tension with *must* ϕ was rated as significantly more acceptable than the matching tension with *certain* ϕ, especially when these are conjoined with *possible* $\neg\phi$. According to the strong view, *must* is at least as strong as *certain*. Suppose we invoke (J-A) to explain why we can assert *must* ϕ when one's evidence allows for the strict possibility that $\neg\phi$. Since norms of assertion are general, we are then also committed to this leniency with respect to assertions of *certain* ϕ, in otherwise matching contexts. As a result, we have no explanation of why tensions with *must* ϕ are nevertheless significantly more acceptable. So adopting (J-A), by itself, does not allow us to reconcile strong *must* with the overall pattern observed in Experiments 1A-B.[26]

Another way of trying to reconcile strong *must* with our results is to reject the assumption that, in the configurations and contexts used in our experiments, *must* and *certain* are assigned modal bases that are equally flexible. Specifically, one could hold that, compared to *must*, the modal base for *certain* is constrained to be more egocentric (cf. von Fintel and Gillies, 2016), which in our (unembedded) examples means that it is constrained to be more speaker-centered. From this perspective, the modal base for *must*/*might* is still determined by f_e, but the modal base for *certain* is determined by f_{e+s}, where the latter function is constrained to represent knowledge states that include that of the subject. The additional flexibility of *must* over *certain* with respect to the modal base could then be used to explain why the former tends to be more acceptable in (matching) contexts of epistemic tension.

Interestingly, however, Experiment 1A has one condition which includes unambiguously speaker-centric epistemic operators, namely, the tensions with *I don't know for sure that* ϕ. This proposal arguably predicts that tensions of the form *certain* $\phi \wedge$ *I don't know for sure* ϕ should be *less* acceptable than those of the form *certain* $\phi \wedge$ *might* $\neg\phi$. For the former would mean something like 'I am

[25] As a full defense of strong *must* + weak *might* package in (5), this response is quite limited. For it doesn't issue in an explanation for why, assuming that *might* and *possible* are just bare possibility operators, *might* behaves as if it is stronger than *possible*.

[26] In other words, we are *not* rejecting (J-A). The point is just that, in contexts of epistemic tensions, *certain* behaves like a stronger epistemic operator than *must*, suggesting that the latter is not maximally strong independently of which general norm of assertion one assumes.

certain that ϕ but I am not sure that ϕ', which feels odd and incoherent, and would be hard to rescue due to the (hypothesized) speaker-centric constrain on the modal base. In contrast, due to the (hypothesized) additional flexibility in the epistemic modal bases that can be used by *might*, tensions of the form *certain* ϕ $\wedge$ *might* $\neg\phi$ could be read as 'I am certain that ϕ but modal base B is compatible with not ϕ', which is coherent, even if somewhat odd. However, this prediction is incorrect: in both the standard and reverse order of the conjuncts, tensions of the form *certain* $\phi \wedge$ *I don't know for sure* ϕ are judged as more acceptable than *certain* $\phi \wedge$ *might* $\neg\phi$ (see Table 2). Hence in the contexts and configurations used in our Experiments, *certain*/*possible* are unlikely to be more speaker-centric and less flexible than *must*/*might*.[27]

To be sure, one could still defend the strong *must* + weak *might* package by trying various combinations of the strategies just discussed. One route is to explore whether there is a systematic difference in the way that epistemic auxiliaries and adjectives interact with slack regulators (but see §2.1.3), or with norms of assertion, that can help this view make sense of our overall pattern of results.[28]

4.2 Revisiting the probabilistic weak *must* + strong *might* package

The results of Experiments 2A-B are in tension with the probabilistic weak *must* + strong *might* account. To be sure, Lassiter (2016)'s official probabilistic account is more complex than the one in (6). Still, most of his additional machinery targets the indirectness evidential component. As a result, his account does not make different predictions in Experiments 2A-B: for in each vignette the evidence for the prejacent of the original modally hedged claims was intentionally designed to be indirect. Indeed, it is useful to recall Lassiter (2016)'s response to von Fintel and Gillies (2010)'s claim that downplaying previous *must*-claims feels substantially worse than downplaying previous claims of very high but non-maximal certainty. Lassiter rejects this intuition and claims that both kinds of downplaying responses feel equally unjustified. However, Experiments 2A-B support von Fintel and Gillies's intuition.

Can we modify the probabilistic account so as to capture our results? The source of the tension with Experiments 2A-B is that treating *must* as expressing high but non-maximal certainty over an epistemic modal base predicts acceptability judgments in downplaying contexts that are too lenient. Specifically, it incorrectly predicts that downplaying *must*-claims should be as acceptable as downplaying other claims which express high but non-maximal certainty. In light of this, consider the tempting revision in (28) (focusing for brevity just on *must*). Instead of relativizing to a pure epistemic space (accessible from w via e), we

[27] For evidence that (un-embedded) *must*/*might* are by default speaker centric, see Roberts (2015).

[28] This way of defending the strong *must* + weak *might* account faces an important obstacle. For simplicity, let us focus on the slack strategy. To explain the result that tensions with *must* are rated as more acceptable than matching tensions with *certain*, we would have to say that *must* allows more slack than *certain*. However, to explain the result that tensions with *might* are less acceptable than matching tensions with *possible*, we would have to say that *might* allows less slack than *possible*. It seems hard to come up with a principled account of these piecemeal differences in tolerance to slack, since they cut across the modal adjectives and auxiliaries.

relativize to an epistemic space which incorporates the (possibly non-veridical) normality assumptions (accessible from w via e^g).

(28) Weak *must* + strong *might* (probabilistic$_g$ version):

a. $[\![must\ \phi]\!]^{w,e^g} = 1$ iff $Pr_{e^g(w)}(\{w' : [\![\phi]\!]^{w',e^g} = 1\}) > \theta_c$

e^g comes from e and a stereotypical/normality conversational background g as follows: if $e(w) = \langle E, Pr\rangle$, then $e^g := \langle E^g, Pr^g\rangle$, where E^g is defined as $\bigcap g(w) \cap E$ and Pr^g is defined as Pr conditionalized on $\bigcap g(w)$.[29] Crucially, this proposal *weakens* the original probabilistic entry for *must* in (6-a). For this reason, it doesn't help with Experiments 2A-B. What we need is a fix in the opposite direction, i.e., one that *strengthens* the original probabilistic entry for weak *must*. In short, given (28), *must* ϕ claims are weaker than on the original probabilistic account, so downplaying *must* claims should be even more acceptable than originally predicted, which was already too lenient.

One option for strengthening *must*—which Lassiter discusses but does not endorse—is to add veridicality to the weak probabilistic account by requiring that the prejacent hold in the world of evaluation, as in (29). This can help explain why it is odd to downplay a previous *must*-claim when the prejacent turns out to be false. Unfortunately, this move negatively affects an advantage of the original account: namely, the strict compatibility of *must* ϕ with the bare possibility that $\neg\phi$. As a result, (29) improves the predictions of the weak probabilistic account vis-á-vis Experiments 2A-B, but downgrades the original predictions vis-á-vis Experiments 1A-B.

(29) Weak *must* + strong *might* (probabilistic$_v$ version):

a. $[\![must\ \phi]\!]^{w,e} = 1$ iff $[\![\phi]\!]^{w,e} = 1\ \wedge\ Pr_{e(w)}(\{w' : [\![\phi]\!]^{w',e} = 1\}) > \theta_c$

To be sure, one could try to reconcile (29-a) with the results of Experiments 1A-B by adopting a weak norm of assertion, such as (J-A). As discussed in §4.1, however, the challenge is then to explain why, given a weak norm of assertion, veridical *must*-claims are more tolerant to conflicting possibilities than matching *certain*-claims. It is an open question whether this challenge can be met in a principled way.[30]

A different way of trying to strengthening *must* is the following. Maintain the basic format of the weak probabilistic account, including the pure epistemic

[29] This is basically a simple adaption of Yalcin (2010)'s probabilistic modification of Kratzer's semantics for conditionals. The difference is that what we are incorporating into the probabilistic space is not an overt antecedent but a set of normality assumptions about the world, as determined by g.

[30] One way of trying to save something like (29), suggested to us by Swanson (pc.), is to revise (29-a) so that the prejacent is pragmatically presupposed (instead of encoding it as either a semantic presupposition or as part of the assertoric content of *must*). One could then appeal to general principles and differences in the corresponding tasks/contexts to try to explain why *must* doesn't (tend to) presuppose the prejacent in tensions like those in Experiments 1A-B, but (tends to) do so in downplaying tensions like those in Experiments 2A-B. This option is worth exploring in detail, but it faces an obstacle. Consider Table 1 in Experiment 1A: note that all the epistemic tensions in the 'standard order'—i.e., when *must* ϕ comes first—are rated better than in the 'reverse order'—i.e., when *must* ϕ comes last. According to this suggestion, however, shouldn't we expect the opposite pattern? For if *must* ϕ defeasibly presupposes ϕ, this presupposition should be suspended when we have clear information that $\neg\phi$ is possible, which is precisely the case in the reverse but not in the standard order. Yet epistemic tensions in the standard order received reliably better ratings than in the reverse order.

language hypothesis, but reconfigure the original relations between the thresholds for the relevant modal expressions. The aim is to try to 'fit' the key patterns in Experiments 1A-B and 2A-B. To illustrate this kind of response, let us focus on *must*, *certain*, and *almost certain*, where their respective thresholds are now stipulated to respect the constrain in (30). Crucially, we can enrich this account with a simple 'Lockean' theory of belief according to which $B_S(\phi)$ if the probability of ϕ for S is at or above some threshold which can be less than 1 (cf. Foley, 1992). To be useful in the current dialectic, however, the belief threshold would have to be set as in (30-a). This would help us mimic the restricted quantificational account of Experiments 2A-B, esp., to explain the substantial difference between tolerance to downplaying in the *almost certain* vs. the *must* conditions. At the same time, *must* ϕ would still be compatible with the bare possibility that $\neg\phi$, thus respecting the patterns in Experiments 1A-B.

(30) Revised constrains on thresholds, for each context c:
a. $\theta_c^{almost\ certain} \leq \theta_c^{believe} \leq \theta^{must} \leq \theta_c^{certain}$

Although this revised probabilistic account seems to fit the results of our experiments (unsurprisingly since it is essentially a post hoc fix which our studies are not designed to test), it faces other problems. First, the stimuli and results of Experiments 2A-B suggest that the probabilistic threshold for *must*, θ^{must}, would in general have to be oddly high (for recall that it is significantly easier to downplay 'n% certain ϕ' than *must* ϕ for n quite close to 100). A similar problem can be raised about the stipulated threshold for belief. It seems quite easy to think of scenarios where one can believe that ϕ, and assert *must* ϕ, even if one is not committed to also accepting that the probability of ϕ, in the relevant pure epistemic space, is say greater than 98%. At any rate, this account could be empirically explored in future work, should theorists still find it prima facie reasonable.

To be clear, we are not rejecting probabilistic accounts of *must* and *might* as such. Our point is just that contexts of epistemic tension suggest that we need to abandon accounts which assume that the auxiliaries are sensitive, by default and directly, to just a pure epistemic modal base. The weak probabilistic account in (6) is one way of holding on to the pure epistemic language hypothesis, and this is the commitment that should be abandoned. There might be other reasons—e.g., gradability, context-dependency, embedability, or inferential relations between epistemic terms—to think that the auxiliaries should be given a probabilisitic semantics (cf. Moss, 2015; Swanson, 2006, 2016; Lassiter, 2017). Indeed, our results are compatible with a probabilisitic account along the lines of (31) (cf. Yalcin, 2005; Moss, 2015). On this view, *must* expresses maximal probability, but achieves the required weakness by incorporating normality assumptions. If we assume, as we did in §3, that the normality assumptions used by the auxiliaries are a subset of those used to determine the space that belief quantifies over, then we can also derive the result that *must*-claims have doxastic entailments. As a result, *must* is still a weaker epistemic operator than *certain*, defined as in (4-a), yet has stronger doxastic commitments with respect to corresponding bare prejacent than 'almost certain'. Furthermore, *might* expresses probability > 0 over that same constrained space, and is thus stronger than a bare possibility operator, such as the one in (4-b).

(31) Weak *must* + strong *might* (probabilistic$_g$ version):

a. $[\![must\ \phi]\!]^{w,e^g} = 1$ iff $Pr_{e^g(w)}(\{w' : [\![\phi]\!]^{w',e^g} = 1\}) = 1$

b. $[\![might\ \phi]\!]^{w,e^g} = 1$ iff $Pr_{e^g(w)}(\{w' : [\![\phi]\!]^{w',e^g} = 1\}) > 0$

A probabilistic account like (31), paired with the entries for the adjectives in (4), captures the right strength and logical relations between the auxiliaries and adjectives to explain our pattern of results in Experiments 1A-B and 2A-B. Crucially, however, this account abandons the pure epistemic language hypothesis for *must* and *might*, makes essential use of normality assumptions, and in this sense directly incorporates a key component of Kratzer-style restricted quantificational accounts.

4.3 Revisiting the quantificational weak *must* + strong *might* package

The restricted quantificational account of weak *must* + strong *might* has two components crucial to explain our results (cf. Kratzer, 1981, 2012; Roberts, 2015; Giannakidou and Mari, 2016). First, it abandons the pure epistemic language hypothesis. The auxiliaries are directly sensitive to epistemic spaces structured and constrained by normality assumptions and expectations about the way the world is. Such spaces are not guaranteed to be realistic. Second, it treats *must* as a universal and *might* as an existential quantifier over the structured and constrained epistemic space, while the epistemic adjectives are assumed to operate directly over spaces (closer to) pure epistemic spaces.[31] It follows that *must* ϕ is weaker than *certain* ϕ, and is strictly compatible with *possible* $\neg\phi$, while *might* ϕ is stronger than *possible* ϕ. Importantly, we assumed that the ordering sources used by epistemic *must* and *might* have to satisfy a doxastic constraint (cf. Swanson, 2016). That is, the space accessible to these auxiliaries is constrained by normality assumptions which speakers/subjects believe, relative to particular deliberation contexts. It follows that, typically, an assertion by S of *must* ϕ entails $B_S(\phi)$, even if it does not entail ϕ, or $K_S(\phi)$. From this perspective, a *must*-claim whose prejacent turns out to be false can be deficient in just the way that an incorrect unhedged belief can be deficient. This helps explain why, even if an assertion of *must* ϕ does not entail ϕ, it is nevertheless hard to downplay it when ϕ turned out to be false. We argued that this restricted quantificational account provides principled explanations of the overall pattern of results observed in Experiments 1A-B and 2A-B.

To be clear, we are not suggesting that this account is essentially complete, nor that it is the only defensible version, in light of our results, of a Kratzerian account of the auxiliaries. Indeed, Kratzer's own original formulation of a stereotypical conversational background g_s is 'deliberatively unconstrained' (2012, p. 37). So it is worth exploring whether we can adopt a formulation that relaxes

[31] Different versions of the mantra may take different positions on the semantics for epistemic adjectives. What is important is that they all posit enough structure to distinguish between access to pure and ordered/biased epistemic modal bases, so they can all easily incorporate our background assumptions. Still, we can, at this point, relax the assumption that *certain* and *possible* operate over a pure epistemic space. To respect our pattern of results, we need only hold that, for each context c, *certain* and *possible* operate over a superset of the space used by the epistemic auxiliaries. This could be a pure epistemic space, or one that is constrained by some normality assumptions, as long as the latter is, in general, a subset of the set of normality assumptions used by the auxiliaries.

the doxastic constraint, introduced in (19), and still predict the results of Experiments 2A-B. Take (32), where 'D_S' stands for the speaker (or relevant subject in embedded cases), and consider a function g_{s_1} that satisfies just (32-a). Suppose that a normality assumption is merely 'reasonable' if D_S does not assign to it the status of full belief but only a 'more likely than not' status. Given this stipulation, *must*-claims can be weaker than matching 'almost certain' claims. As a result, we no longer predict that *must*-claims should be harder to downplay than claims of high but non-maximal certainty. For example, in (20), Alex's downplaying claim could be interpreted as insisting that he only asserted that the proposition that it is raining follows given what he thought was relatively reasonable. Intuitively, this seems like the sort of claim that should be easy to downplay.

(32) **Stereotypical conversational background*.** A function g_s such that for any world w, $g_s(w)$ represents what is normal in w according to some *suitable standard* in w.

a. A set of premises P *suitably represents* what is normal in w if it captures the reasonable expectations of D_S concerning w.

Now, one could agree that, to explain our results, we need to impose stronger constraints on g_s, yet still insist that going for a doxastic constraint such as (19) is not the only or best option. For example, Roberts (2015) holds that the ordering source for *must* can include propositions that are 'supposed even if not believed'. The challenge for this proposal, again, is to explain why *must*-claims are stronger, or at least harder to downplay, than claims of very high but non-maximal certainty. Yet consider the proposal in (33-a), where 'D_P' stands for the set of discourse participants and 'D_G' for the interlocutors' mutually evident goals. On this view, a suitable standard for the normality assumptions consists in what interlocutors jointly suppose. Alex's downplaying move would amount to insisting that he only asserted that the proposition that it is raining follows from the evidence and the joint suppositions of the interlocutors, given their goals. This move can be wrong, or unjustified, if the set of joint suppositions were in fact (or could reasonably be inferred to be in our target scenario) weaker than Alex assumed.

(33) **Stereotypical conversational background**.** A function g^*_s such that for any world w, $g^*_s(w)$ represents what is normal in w according to some *suitable standard* for w.

a. A set of premises P *suitably represents* what is normal in w if it captures the joint suppositions, relative to D_G, of D_P concerning w.

To be sure, (33) is just one way to constrain the notion of a stereotypical conversational background, for modal terms such as *must* and *might*, without appealing directly to doxastic constraints on speakers/subjects. Clearly, there is still plenty of empirical and theoretical work to do before we fully understand the nature of normality assumptions, their constraints, and how they and related types of ordering sources are incorporated into the semantics different kinds of modals.[32]

[32] To support the claim that the ordering sources for *must/might* have more to do with reasonable suppositions than normality assumptions that are fully believed, Roberts (2015) points out the intuitive coherence of expressions like *Jane must be the murderer, but I can't believe it.* However, these sorts of examples seem, to us, to have a kind of incapacity reading (as in 'I can't get to myself to accept the consequences of what I strictly have come believe').

Our defense of the restricted quantificational account raises an important foundational question, which we would like to briefly discuss before concluding: Why should natural languages have terms that operate, by default, over epistemic spaces ordered by normality assumptions, rather than just having terms that operate directly over pure epistemic modal bases?

Accounts which accept the pure epistemic language hypothesis for *must* and *might*, such as the strong *must* + weak *might* account of von Fintel and Gillies (2010), are elegant and theoretically simple. However, this theoretical simplicity doesn't directly map unto a simple hypothesis about the interface between these epistemic terms and general cognition. Why? Reasoning from pure epistemic bases is, arguably, an advanced cognitive achievement. In general, we use evidence to make inferences and decisions only against a rich set of normality assumptions and expectations (often implicit and opaque) about the world. In most cases that can be broadly classified as instances of 'common-sense reasoning', we reason as if our world belongs to the 'best' worlds compatible with what we know. This is undoubtedly how infants and non-human animals reason from evidence: they automatically integrate incoming information with a rich set of assumptions and expectations about the way the world is.[33] From this perspective, it makes sense that we should have a set of basic epistemic terms ready made to, by default, latch unto these normality-ordered non-realistic spaces. According to the restricted quantificational account, this is the role of *must* and *might*.

The conjecture that modal terms can be distinguished in terms of the degree to which they use modal spaces continuous with those used by typical instances of common-sense reasoning opens up some interesting future projects. According to the restricted quantificational account we have defended here, epistemic auxiliaries such as *must* and *might* latch unto subsets of the 'best' worlds of epistemic spaces, whereas epistemic adjectives such as *certain*, *95% certain* and *possible* tend to use modal spaces that more closely approximate pure epistemic spaces. Assuming that ordered sets of 'best' epistemic spaces capture our primitive space for modal reasoning, this view suggests that, during development, the auxiliaries should be easier to acquire, and in general less taxing to use, than the adjectives: i.e., epistemic *must* should be easier than *certain*, and *might* easier than *possible*, when controlling for other differences in the acquisition and processing of auxiliaries

Furthermore, expressions like *I checked the weather channel. It must be raining. # Indeed, I believe that it is raining*, feel odd and redundant, as is brought out by comparing that to *I checked the weather channel. It's very likely/almost certain that it is raining. Indeed, I believe it is raining*). This contrast is expected on our account, whereas Roberts predicts that expressions of the form 'must ϕ; indeed, I believe ϕ' should, in general, feel informative. Overall, we are not, at this point, convinced that there is any need to abandon the doxastically constrainted version of stereotypical ordering sources for *must* and *might*. Still, experimental investigations into the relations between epistemic terms and attitudes like *believes* and *knows*, especially using contexts of epistemic tension, are a promising next step in addressing some of these open issues.

[33] There are fascinating debates about the source of these assumptions and expectations about the world, from empiricists who argue that they are acquired via general learning processes (Gopnik, 2003; Gopnik and Schulz, 2004), to core cognition advocates who think that much of this basic framework is innate, i.e., part of the basic cognitive architecture of this or that species (Spelke and Kinzler, 2007; Carey, 2009). Crucially, all these theorists agree that, by the time children start learning language, they have in place a rich set of assumptions and expectations—from our perspective, a rich set of available normality assumptions—that are automatically integrated with evidence to reason in specific domains and contexts.

vs. adjectives.[34] More generally, modal terms which live on ordered spaces should be easier to acquire and use than terms which live on pure epistemic spaces. If this suggestion—at this stage admittedly quite speculative—is roughly correct, we can explore and compare the predictions, for acquisition and processing, of theories of epistemic modals which differ with respect to their classification of which modals live on (approximately) pure epistemic modals bases and which live on (common-sense like) normality-based ordered spaces.[35]

5 Conclusion

Contexts of epistemic tension can be fruitfully used to test different accounts of epistemic terms (cf. Yalcin, 2007, 2010; von Fintel and Gillies, 2010, 2016; Willer, 2013; Roberts, 2015; Swanson, 2016; Lassiter, 2016, 2017; Giannakidou and Mari, 2016; Ninan, 2018). At the same time, the target judgments are often both subtle and under dispute amongst theorists. To make progress in these debates, we should obtain or corroborate these judgments via controlled experimental settings such as those used in Experiments 1 and 2. This approach allowed us, in addition, to generate a substantially greater range of comparative acceptability ratings than has thus far been considered in the literature. This comparative data proved crucial to test the unique predictions of three influential recent accounts of epistemic *must* and *might*, especially when the theories are paired with (plausible, sometimes language independent) additional general principles (e.g., pragmatic slack or mid-discourse change of mind). We argued that the overall pattern of results not only supports the restricted quantificational account of weak *must* + strong *might* (Kratzer, 1981, 1991, 2012; Roberts, 2015; Giannakidou and Mari, 2016), but also help us identify which of their components are responsible for its success.

[34] To illustrate this point, consider a real-world 'disjunction elimination'-like pattern of reasoning. This kind of inference can be easily performed by adults, infants, and some non-human animals (Cesana-Arlotti et al, 2018; Andrews, 2016). (i) S sees that the candy is placed in one of two boxes, A or B, but doesn't know in which one (they were shuffled). (ii) Box A is then opened and S can then see that the candy is not there. (iii) S then concludes that the candy is in box B. The transition from (i) to (iii), when it's part of temporally extended sequences of events, is made against rich background assumptions about object individuation and other principles of naive physics. It would be harder for S to instead reason from pure modal bases and conclude that, e.g., since some principles of object individuation could cease to hold between events (ii) and (iii), it is strictly possible that the candy is not in B (even if it is still likely it is in B). To be sure, approximations to the latter kind of purely evidence-based reasoning are clearly within the capacity of adults, and they can be triggered in various ways (e.g., with linguistic cues). Still, they are importantly different from (standard cases of) common-sense reasoning—specifically, they require that we separate, to some extent, the objective evidence from various background normality assumptions, a process that is presumably not easy to execute. Against this background, suppose that Mhas to evaluate an assertion by S, after stages (i)-(ii), of *the candy must be in B* vs. *it's certain that the candy is in B*. A theory which entails that *must* uses ordered epistemic spaces, whereas *might* uses pure ones, predicts that verifying the latter claim should be harder. This is because it requires that we suspend or take a step back from (to some extent at least) the common-sense processes that would usually/naturally lead us to a conclusion analogous to (iii).

[35] For example, according to the account of von Fintel and Gillies (2010), *must*, which is hypothesized to be a pure epistemic term, should be harder to learn and more taxing to use than epistemic *should*, which is also universal but operates on normality-ordered non-realistic spaces.

Appendix A Stimuli for Experiment 1A

(1) I just bought a vintage bicycle at a garage sale in my neighborhood. It will need some work, but it's in decent shape. [**EPISTEMIC TENSION HERE**] The previous owner didn't even know the name of the manufacturer

a. Tensions with *must*: presented in canonical order (shown here) AND in reverse order (not shown):
   (i) The bike must be from the 1960s, but I don't know for sure.
   (ii) The bike must be from the 1960s, but its possible that it isn't.
   (iii) The bike must be from the 1960s, but it might not be.
   (iv) The bike must be from the 1960s, but it isn't.

b. Tensions with *certain*: presented in canonical order (shown here) AND in reverse order (not shown):
   (i) It's certain that the bike is from the 1960s, but I don't know for sure.
   (ii) It's certain that the bike is from the 1960s, but it's possible that it isn't.
   (iii) It's certain that the bike is from the 1960s, but it might not be.
   (iv) It's certain that the bike is from the 1960s, but it isn't.

(2) When football teams buy helmets, they rarely repaint them in the team's colors. [**EPISTEMIC TENSION HERE**] I suppose you don't want to apply coats of paint to the helmet just to have the paint chip off during gameplay.

a. Tensions with *must*: presented in canonical order (shown here) AND in reverse order (not shown):
   (i) Repainting them must be too much of a hassle, but I don't know for sure.
   (ii) Repainting them must be too much of a hassle, but it's possible that it isn't.
   (iii) Repainting them must be too much of a hassle, but it might not be.
   (iv) Repainting them must be too much of a hassle, but it isn't.

b. Tensions with *certain*: presented in canonical order (shown here) AND in reverse order (not shown):
   (i) It's certain that repainting them is too much of a hassle, but I don't know for sure.
   (ii) It's certain that repainting them is too much of a hassle, but its possible that it isn't.
   (iii) It's certain that repainting them is too much of a hassle, but it might not be.
   (iv) It's certain that repainting them is too much of a hassle, but it isn't.

(3) I think I've found my dream car at a used car dealership down the road: a beautiful 1964 white Ford Mustang. The body, paint, and suede interior look pristine. I checked the speedometer, and it shows 38,000 miles. [**EPISTEMIC TENSION HERE**] At any rate, the car drove beautifully during the test drive!

a. Tensions with *must*: presented in canonical order (shown here) AND in reverse order (not shown):
   (i) The mileage must actually be 138,000, but I don't know for sure.
   (ii) The mileage must actually be 138,000, but its possible that it isn't.
   (iii) The mileage must actually be 138,000, but it might not be.
   (iv) The mileage must actually be 138,000, but it isn't.

b. Tensions with *certain*: presented in canonical order (shown here) AND in reverse order (not shown):
   (i) It's certain that the mileage is actually 138,000, but I don't know for sure.
   (ii) It's certain that the mileage is actually 138,000, but its possible that it isn't.
   (iii) It's certain that the mileage is actually 138,000, but it might not be
   (iv) It's certain that the mileage is actually 138,000, it isn't.

(4) Jack Spencer used to produce his photographs in a darkroom. Lately, though, he's been working with inkjet printing. [**EPISTEMIC TENSION HERE**] Either way, he is an incredibly gifted photographer, and I'm a big fan of his work.

a. Tensions with *must*: presented in canonical order (shown here) AND in reverse order (not shown):
   (i) So he must be working with Photoshop as well, but I don't know for sure.
   (ii) So he must be working with Photoshop as well, but its possible that he isn't.
   (iii) So he must be working with Photoshop as well, but he might not be.
   (iv) So he must be working with Photoshop as well, but he isn't.

b. Tensions with *certain*: presented in canonical order (shown here) AND in reverse order (not shown):
   (i) So it's certain that he's working with Photoshop as well, but I don't know for sure.
   (ii) So it's certain that he's working with Photoshop as well, but its possible that he isn't.

(iii) So it's certain that he's working with Photoshop as well, but he might not be.
(iv) So it's certain that he's working with Photoshop as well, but he isn't.

(5) It is getting more difficult to find stereo systems that will play cassettes. You could try a used electronics store, or if you want to buy a brand new stereo, the Jensen JTA has a built-in cassette deck. [**EPISTEMIC TENSION HERE**]

a. Tensions with *must*: presented in canonical order (shown here) AND in reverse order (not shown):
(i) There must be others on the market, but I don't know for sure.
(ii) There must be others on the market, but its possible that there aren't.
(iii) There must be others on the market, but there might not be.
(iv) There must be others on the market, but there aren't.

b. Tensions with *certain*: presented in canonical order (shown here) AND in reverse order (not shown):
(i) It's certain that there are others on the market, but I don't know for sure.
(ii) It's certain that there are others on the market, but its possible that there aren't.
(iii) It's certain that there are others on the market, but there might not be.
(iv) It's certain that there are others on the market, but there aren't.

Appendix B Stimuli for Experiment 1B

(1) I just bought a vintage bicycle at a garage sale in my neighborhood. It will need some work, but it's in decent shape. [**EPISTEMIC TENSION HERE**]

a. *must* and *certain* + neutral conjuncts:
(i) The bike must be from the 1960s, but I don't know where it was produced.
(ii) It's certain that the bike is from the 1960s, but I don't know where it was produced.

b. *must* and *certain* in epistemic tensions:
(i) The bike must be from the 1960s, but its possible that it isn't.
(ii) It's certain that the bike is from the 1960s, but its possible that it isn't.

(2) When football teams buy helmets, they rarely repaint them in the team's colors. [**EPISTEMIC TENSION HERE**]

a. *must* and *certain* + neutral conjuncts:
(i) Repainting them must be too much of a hassle, but I don't think that the teams even care whether the colors match the uniforms.
(ii) It's certain that repainting them is too much of a hassle, but I don't think that the teams even care whether the colors match the uniforms.

b. *must* and *certain* in epistemic tensions:
(i) Repainting them must be too much of a hassle, but it's possible that it isn't.
(ii) It's certain that repainting them is too much of a hassle, but its possible that it isn't.

(3) I think I've found my dream car at a used car dealership down the road: a beautiful 1964 white Ford Mustang. The body, paint, and suede interior look pristine. I checked the speedometer, and it shows 38,000 miles. [**EPISTEMIC TENSION HERE**]

a. *must* and *certain* + neutral conjuncts:
(i) The mileage must actually be 138,000, but I don't think this will affect the car's performance.
(ii) It's certain that the mileage is actually be 138,000, but I don't think this will affect the car's performance.

b. *must* and *certain* in epistemic tensions:
(i) The mileage must actually be 138,000, but its possible that it isn't.
(ii) It's certain that the mileage is actually 138,000, but its possible that it isn't.

(4) Jack Spencer used to produce his photographs in a darkroom. Lately, though, he's been working with inkjet printing. [**EPISTEMIC TENSION HERE**]

a. *must* and *certain* + neutral conjuncts:
(i) So he must be working with Photoshop as well, but I don't think he has stopped working in the darkroom altogether.
(ii) So it's certain that he's working with Photoshop as well, but I don't think he has stopped working in the darkroom altogether.

b. *must* and *certain* in epistemic tensions:
(i) So he must be working with Photoshop as well, but its possible that he isn't.
(ii) So it's certain that he's working with Photoshop as well, but its possible that he isn't.

(5) It is getting more difficult to find stereo systems that will play cassettes. You could try a used electronics store, or if you want to buy a brand new stereo, the Jensen JTA has a built-in cassette deck. [**EPISTEMIC TENSION HERE**]

a. *must* and *certain* + neutral conjuncts:
   (i) There must be others on the market as well, but I don't know if they are affordable
   (ii) It's certain that there are others on the market as well, but I don't know if they are affordable.

b. *must* and *certain* in epistemic tensions:
   (i) There must be others on the market, but its possible that there aren't.
   (ii) It's certain that there are others on the market, but its possible that there aren't.

Appendix C Stimuli for Experiment 2A

(1) *Alex and Billy are deliberating about whether to go outside for a moment. They will only do so if it is not raining. Alex looks at two weather forecasting websites, and both say it is raining at their location.*

Alex: It must be raining outside. / It is 95% certain that it's raining outside. / It is certain that it's raining outside.
Billy opens the curtains. He and Alex can both see that it is actually not raining outside.
Billy: No it isn't. You were wrong.
Alex: [**DOWNPLAYING STATEMENT GOES HERE**] Anyway, let's go outside then.

a. Downplaying conditions (presented in **BOLD**):
   (i) Well, strictly speaking, I was not wrong. I only said that it must be raining outside.
   (ii) Well, strictly speaking, I was not wrong. I only said that it is 95% certain that it's raining outside.
   (iii) Well, strictly speaking, I was not wrong. I only said that it is certain that it's raining outside.

(2) *Detectives Billy and Maria are deliberating about whether Smith murdered Jacob. They are required to file their final report to the court. Billy evaluates the interviews of several independent eyewitnesses at the scene. All say that Smith was the murderer.*

Billy: All the eye witnesses agree. Smith must be the murderer. / It is almost certain that Smith is the murderer. / It is certain that Smith is the murderer
Maria hands Billy the latest DNA report, which shows that Smith's DNA was nowhere to be found at the scene.
Maria: You were wrong. Smith wasn't even at the scene of the crime.
Alex: [**DOWNPLAYING STATEMENT GOES HERE**] Anyway, we'll have to tell the court that it couldn't have been Smith.

a. Downplaying conditions (presented in **BOLD**):
   (i) Well, strictly speaking, I was not wrong. I only said that Smith must be the murderer.
   (ii) Well, strictly speaking, I was not wrong. I only said that it is almost certain that Smith was the murderer.
   (iii) Well, strictly speaking, I was not wrong. I only said that it is certain that Smith was the murderer.

(3) *Officers Wolfgang and Dressler are deliberating whether Amy Hauser's mother, Joanne Hauser, was born in Germany before 1960. If she was, then Amy Hauser has the right to German citizenship. As part of the application, a birth certificate was submitted which says that Joanne Hauser was born in Germany in 1958.*

Wolfgang: The certificate is here. Joanne Hauser must have been born in Germany in 1958. / It is almost certain that Joanne Hauser was born in Germany in 1958. / It is certain that Joanne Hauser was born in Germany in 1958.
Dressler opens the results of the document authenticity checks, and they show that Joanne Hauser's birth certificate was forged.
Dressler: No she wasn't. You were wrong.
Wolfgang: [**DOWNPLAYING STATEMENT GOES HERE**] Anyway, let's reject Amy's citizenship claim then.

a. Downplaying conditions (presented in **BOLD**):

(i) Well, strictly speaking, I was not wrong. I only said that Joanne Hauser must have been born in Germany in 1958.
(ii) Well, strictly speaking, I was not wrong. I only said that it is almost certain that Joanne Hauser was born in Germany in 1958.
(iii) Well, strictly speaking, I was not wrong. I only said that it is certain that Joanne Hauser was born in Germany in 1958.

(4) *Jane and Moira are the managers of the supermarket where Gary works. Today, Gary has not shown up for work. Jane and Moira are deliberating about whether to punish Gary, possibly even by firing him. Moira just saw on Facebook that someone uploaded a picture of Gary at the beach.*

Moira: Look at the picture! Gary must have skipped work for fun today. / It is almost certain that Gary skipped work for fun today. / It is certain that Gary skipped work for fun today.
Jane hangs up her phone. Someone from hospital has just informed her that Gary is currently being treated for injuries.
Jane: That was the hospital: Gary was in a car accident yesterday. So you were wrong.
Moira: [**DOWNPLAYING STATEMENT GOES HERE**] Anyway, let's go to the hospital later to make sure he is ok.

a. Downplaying conditions (presented in **BOLD**):
(i) Well, strictly speaking, I wasn't wrong. I only said that that Gary must have skipped work for fun today.
(ii) Well, strictly speaking, I wasn't wrong. I only said that it is almost certain that Gary skipped work for fun today.
(iii) Well, strictly speaking, I wasn't wrong. I only said that it is certain that Gary skipped work for fun today.

(5) *Bill and David are at a used car dealership and are deliberating about whether to buy a red sedan in the lot. They will only do so if the car has never been in a major accident. Bill, who is very knowledgable about cars, does a comprehensive inspection of the engine, as well as of the car's interior and exterior. To Bill, the car looks pristine.*

Bill: This car must have never been in an accident. / It is 95% certain that this car has never been in an accident. / It is certain that this car has never been in an accident.
From his phone, David runs an accident history report on the car. According to the report, the car was once repaired after a high-speed collision.
David: Actually, you were wrong. This car was in a major collision two years ago.
Bill: [**DOWNPLAYING STATEMENT GOES HERE**] Anyway, we definitely shouldn't buy this one.

a. Downplaying conditions (presented in **BOLD**):
(i) Well, strictly speaking, I was not wrong. I only said that this car must have never been in an accident.
(ii) Well, strictly speaking, I was not wrong. I only said that it is 95% certain that this car has never been in an accident.
(iii) Well, strictly speaking, I was not wrong. I only said that it is certain that this car has never been in an accident.

Appendix D Stimuli for Experiment 2B

(1) *Alex and Billy are deliberating about whether to go outside for a moment. They will only do so if it is not raining. Alex looks at two weather forecasting websites, and both say it is raining at their location.*

Alex: It must be raining outside. / It is 95% certain that it's raining outside. / It is at least 95% certain that it's raining outside. / It is certain that it's raining outside.
Billy opens the curtains. He and Alex can both see that it is actually not raining outside.
Billy: No it isn't. You were wrong.
Alex: [**DOWNPLAYING STATEMENT GOES HERE**] Anyway, let's go outside then.

a. Downplaying conditions (presented in **BOLD**):
(i) Well, strictly speaking, I was not wrong. I didn't say that it's raining outside. I only said that it must be raining outside..
(ii) Well, strictly speaking, I was not wrong. I didn't say that it's raining outside. I only said that it is 95% certain that it's raining outside.

(iii) Well, strictly speaking, I was not wrong. I didn't say that it's raining outside. I only said that it is at least 95% certain that it's raining outside.
(iv) Well, strictly speaking, I was not wrong. I didn't say that it's raining outside. I only said that it is certain that it's raining outside.

(2) *Detectives Billy and Maria are deliberating about whether Smith murdered Jacob. They are required to file their final report to the court. Billy evaluates the interviews of several independent eyewitnesses at the scene. All say that Smith was the murderer.*

Billy: All the eye witnesses agree. Smith must be the murderer. / It is almost certain that Smith is the murderer. / It is certain that Smith is the murderer
Maria hands Billy the latest DNA report, which shows that Smith's DNA was nowhere to be found at the scene.
Maria: You were wrong. Smith wasn't even at the scene of the crime.
Alex: [**DOWNPLAYING STATEMENT GOES HERE**] Anyway, we'll have to tell the court that it couldn't have been Smith.

a. Downplaying conditions (presented in **BOLD**):
(i) Well, strictly speaking, I was not wrong. I didn't say that Smith is the murderer. I only said that Smith must be the murderer.
(ii) Well, strictly speaking, I was not wrong. I didn't say that Smith is the murderer. I only said that it is 95% certain that Smith is the murderer.
(iii) Well, strictly speaking, I was not wrong. I didn't say that Smith is the murderer. I only said that it is at least 95% certain that Smith is the murderer.
(iv) Well, strictly speaking, I was not wrong. I didn't say that Smith is the murderer. I only said that it is certain that Smith is the murderer.

(3) *Officers Wolfgang and Dressler are deliberating whether Amy Hauser's mother, Joanne Hauser, was born in Germany before 1960. If she was, then Amy Hauser has the right to German citizenship. As part of the application, a birth certificate was submitted which says that Joanne Hauser was born in Germany in 1958.*

Wolfgang: The certificate is here. Joanne Hauser must have been born in Germany in 1958. / It is 95% certain that Joanne Hauser was born in Germany in 1958. / It is at least 95% certain that Joanne Hauser was born in Germany in 1958. / It is certain that Joanne Hauser was born in Germany in 1958.
Dressler opens the results of the document authenticity checks, and they show that Joanne Hauser's birth certificate was forged.
Dressler: No she wasn't. You were wrong.
Wolfgang: [**DOWNPLAYING STATEMENT GOES HERE**] Anyway, let's reject Amy's citizenship claim then.

a. Downplaying conditions (presented in **BOLD**):
(i) Well, strictly speaking, I was not wrong. I didn't that Joanne Hauser was born in German in 1958. I only said that Joanne Hauser must have been born in Germany in 1958.
(ii) Well, strictly speaking, I was not wrong. I didn't that Joanne Hauser was born in German in 1958. I only said that it is 95% certain that Joanne Hauser was born in Germany in 1958.
(iii) Well, strictly speaking, I was not wrong. I didn't that Joanne Hauser was born in German in 1958. I only said that it is at least 95% certain that Joanne Hauser was born in Germany in 1958.
(iv) Well, strictly speaking, I was not wrong. I didn't that Joanne Hauser was born in German in 1958. I only said that it is certain that Joanne Hauser was born in Germany in 1958.

(4) *Jane and Moira are the managers of the supermarket where Gary works. Today, Gary has not shown up for work. Jane and Moira are deliberating about whether to punish Gary, possibly even by firing him. Moira just saw on Facebook that someone uploaded a picture of Gary at the beach.*

Moira: Look at the picture! Gary must have skipped work for fun today. / It is 95% certain that Gary skipped work for fun today. / It is at least 95% certain that Gary skipped work for fun today. / It is certain that Gary skipped work for fun today.
Jane hangs up her phone. Someone from hospital has just informed her that Gary is currently being treated for injuries.
Jane: That was the hospital: Gary was in a car accident yesterday. So you were wrong.
Moira: [**DOWNPLAYING STATEMENT GOES HERE**] Anyway, let's go to the hospital later to make sure he is ok.

a. Downplaying conditions (presented in **BOLD**):

(i) Well, strictly speaking, I wasn't wrong. I didn't say that Gary skipped work for fun today. I only said that Gary must have skipped work for fun today.
(ii) Well, strictly speaking, I wasn't wrong. I didn't say that Gary skipped work for fun today. I only said that it is 95% certain that Gary skipped work for fun today.
(iii) Well, strictly speaking, I wasn't wrong. I didn't say that Gary skipped work for fun today. I only said that it is at least 95% certain that Gary skipped work for fun today.
(iv) Well, strictly speaking, I wasn't wrong. I didn't say that Gary skipped work for fun today. I only said that it is certain that Gary skipped work for fun today.

(5) *Bill and David are at a used car dealership and are deliberating about whether to buy a red sedan in the lot. They will only do so if the car has never been in a major accident. Bill, who is very knowledgable about cars, does a comprehensive inspection of the engine, as well as of the car's interior and exterior. To Bill, the car looks pristine.*

Bill: This car must have never been in an accident. / It is 95% certain that this car has never been in an accident. / It is at least 95% certain that this car has never been in an accident. / It is certain that this car has never been in an accident.
From his phone, David runs an accident history report on the car. According to the report, the car was once repaired after a high-speed collision.
David: Actually, you were wrong. This car was in a major collision two years ago.
Bill: [**DOWNPLAYING STATEMENT GOES HERE**] Anyway, we definitely shouldn't buy this one.

a. Downplaying conditions (presented in **BOLD**):
(i) Well, strictly speaking, I was not wrong. I didn't say that this car has never been in an accident. I only said this car must never have been in an accident.
(ii) Well, strictly speaking, I was not wrong. I didn't say that this car has never been in an accident. I only said that it is 95% certain that this car has never been in an accident.
(iii) Well, strictly speaking, I was not wrong. I didn't say that this car has never been in an accident. I only said that it is at least 95% certain that this car has never been in an accident.
(iv) Well, strictly speaking, I was not wrong. I didn't say that this car has never been in an accident. I only said that it is certain that this car has never been in an accident.

Appendix E Figures

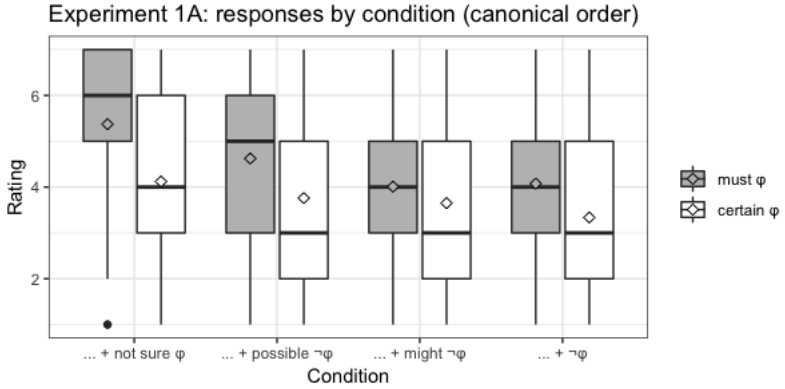

Figure 1A. Box and whiskers plot for Experiment 1A, standard order. The box contains values between the lower and upper quartiles, the horizontal line denotes the median, and the diamond denotes the mean. Whiskers extend to 1.5 times the inter-quartile range (hence, outliers—denoted by black dots—fall out of this range).

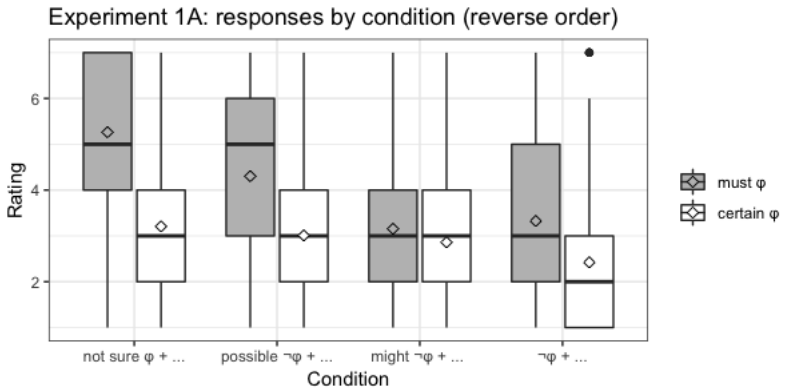

Figure 1A′. Box and whiskers plot for Experiment 1A, reserve order. The box contains values between the lower and upper quartiles, the horizontal line denotes the median, and the diamond denotes the mean. Whiskers extend to 1.5 times the inter-quartile range (hence, outliers—denoted by black dots—fall out of this range).

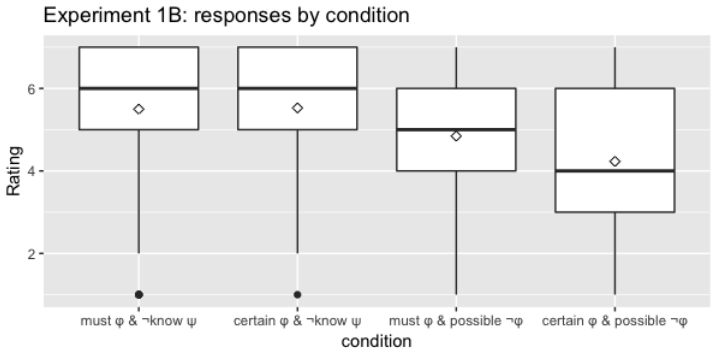

Figure 1B. Box and whiskers plot for Experiment 1B. The box contains values between the lower and upper quartiles, the horizontal line denotes the median, and the diamond denotes the mean. Whiskers extend to 1.5 times the inter-quartile range (hence, outliers—denoted by black dots—fall out of this range).

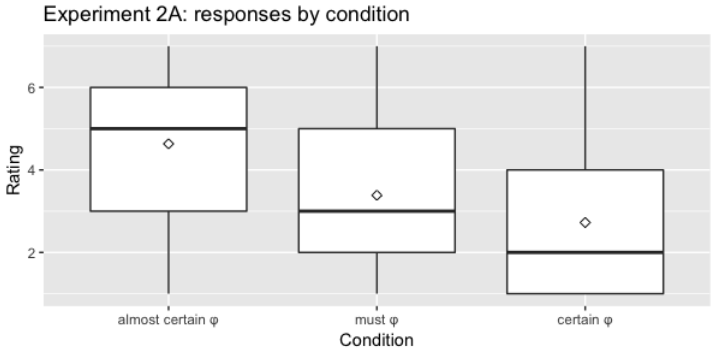

Figure 2A. Box and whiskers plot for Experiment 2A. The box contains values between the lower and upper quartiles, the horizontal line denotes the median, and the diamond denotes the mean. Whiskers extend to 1.5 times the inter-quartile range (hence, outliers—denoted by black dots—fall out of this range).

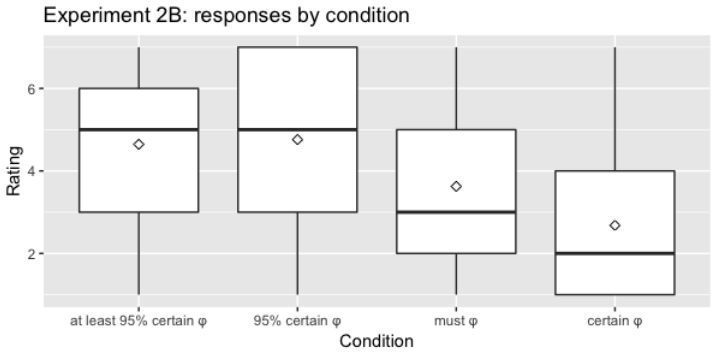

Figure 2B. Box and whiskers plot for Experiment 2B. The box contains values between the lower and upper quartiles, the horizontal line denotes the median, and the diamond denotes the mean. Whiskers extend to 1.5 times the inter-quartile range (hence, outliers—denoted by black dots—fall out of this range).

Acknowledgements Versions of this paper where presented at ZAS Berlin, the Free University of Berlin, University of Michigan, Ann Arbor, Stanford University, University of California, San Diego, and the University of Hamburg. Thanks to audiences at those talks for excellent comments. We would also like to give special thanks to Craige Roberts, Anastacia Giannakidou, Uli Sauerland, Hubert Truckenbrodt, Stephanie Zolt, Julia Zakkou, Brian H. Kim, Eric Swanson, Daniel Lassiter, Marie-Christine Meyer, Ezra Keshet, and Patrick Elliot. Finally, we are also grateful to two anonymous reviewers for NALS who provided us with extremely helpful comments and led to considerable improvements to the manuscript.

References

Alxatib S, Pagin P, Sauerland U (2013) Acceptable contradictions: Pragmatics or semantics? Journal of Philosophical Logic 42:619–634

Andrews K (2016) Animal cognition. In: Zalta ED (ed) Stanford Encyclopedia of Philosophy, Center for the Study of Language and Information, DOI ⟨https://plato.stanford.edu/archives/sum2016/entries/cognition-animal/⟩

Carey S (2009) The Origin of Concepts. Oxford University Press, Oxford

Cesana-Arlotti N, Martín A, Téglás E, Vorobyova L, Cetnarski R, Bonatti LL (2018) Precursors of logical reasoning in preverbal human infants. Science 359(6381):1263–1266

Foley R (1992) The epistemology of belief and the epistemology of degrees of belief. American Philosophical Quarterly 29(2):111–124

Giannakidou A, Mari A (2016) Epistemic future and epistemic *must*: Nonveridicality, evidence and partial knowledge. In: Blaszczak J, Giannakidou A, Klimek-Jankowska D, Migdalski K (eds) Mood, aspect and modality: New answers to old questions, University of Chicago Press, Chicago, IL, chap 3, pp 75–117

Goodhue D (2017) Must p is felicitious only if p is no known. Semantics and Pragmatics 10(14)

Gopnik A (2003) The theory theory as an alternative to the innateness hypothesis. Chomsky and his critics pp 238–254

Gopnik A, Schulz L (2004) Mechanisms of theory formation in young children. Trends in cognitive sciences 8(8):371–377

Hacquard V (2011) Modality. In: Maienborn C, von Heusinger K, Portner P (eds) Semantics: An International Handbook of Natural Language Meaning, vol 2, Mouton de Gruyter, chap 59, pp 1484–1515
Harman G (1986) Change in view. The MIT Press, Cambridge, MA
Khoo J (2014) Modal disagreements. Inquiry 58(5):511–534
Klecha P (2012) Positive and conditional semantics for gradable modals. In: et al AG (ed) Proceedings of Sinn & Bedeutung 16, MIT Working Papers in Linguistics, pp 363–376
Klecha P (2014) Bridging the divide: Scalarity and modality. PhD thesis, University of Chicago
Klecha P (2018) On unidirectionality in precisification. Linguist and Philosophy 41:87–124
Knobe J, Yalcin S (2014) Epistemic modals and context: Experimental data. Semantics and Pragmatics 7:1–21
Kratzer A (1981) The notional category of modality. In: Eikmeyer H, Reiser H (eds) Words, worlds, and contexts: New approaches in word semantics, De Gruyter, Berlin, pp 38–74
Kratzer A (1991) Modality. In: von Stechow A, Wunderlich D (eds) Semantics: An international handbook of contemporary research, De Gruyter, Berlin, pp 639–650
Kratzer A (2012) Modals and Conditionals. Oxford University Press, Oxford
Krifka M (1999) At least some determiners aren't determiners. In: Turner K (ed) The semantics/pragmatics interface from different points of view, vol 1, Elsevier Science Publishers, pp 257–291
Kvanvig JL (2009) Assertion, knowledge, lotteries. In: Greenough P, Pritchard D (eds) Williamson on knowledge, Oxford University Press, Oxford, pp 140–160
Lackey J (2007) Norms of assertion. Noûs 41(4):594–626
Lasersohn P (1999) Pragmatic halos. Language 75(3):522–551
Lassiter D (2016) Must, knowledge, and (in) directness. Natural Language Semantics 24(2):117–163
Lassiter D (2017) Graded Modality: Qualitative and Quantitative Perspectives. Oxford University Press, Oxford, UK
Lauer S (2013) Towards a dynamic pragmatics. PhD thesis, Stanford University
Mandelkern M (2016) A solution to karttunen's problem. In: Sinn und Bedeutung (SuB), vol 21, pp 1–18
Mayr C (2013) Implicatures of modified numerals. In: Caponigro I, Cecchetto C (eds) From Grammar to Meaning: The Spontaneous Logicality of Language, Cambridge University Press, chap 6, pp 139–171
Moss S (2015) On the semantics and pragmatics of epistemic vocabulary. Semantics and Pragmatics 1(1-43)
Ninan D (2018) Relational semantics and domain semantics for epistemic modals. Journal of Philosophical Logic 47(1):1–16
Portner P (2009) Modality. Oxford University Press, Oxford, UK
Roberts C (2015) The character of epistemic modality: Evidentiality, indexicality, and what's at issue, manuscript
Rooth M (1992) A theory of focus interpretation. Natural language semantics 1(1):75–116
Rooth M (1996) Focus. In: ed by Shalom Lappin (ed) The handbook of contemporary semantic theory, Oxford: Blackwell, chap 10, pp 271–97

Santorio P, Romoli J (2017) Probability and implicatures: A unified acccount of the scalar effects of disjunction under modals. Semantics and Pragmatics 10(13)
Sauerland U (2011) Vagueness in language: The case against fuzzy logic revisited. In: Cintula P, Fermüller C, Godo L, Hájek P (eds) Understanding Vagueness: Logical, Philosophical and Linguistic Perspectives, College Publications, London, UK, pp 185–98
Sharvit Y (2017) A note on (strawson) entailment. Semantics and Pragmatics 10:1–38
Spelke ES, Kinzler KD (2007) Core knowledge. Developmental science 10(1):89–96
Swanson E (2006) Interactions with context. PhD thesis, MIT, Cambridge, MA
Swanson E (2011) How not to theorize about the language of subjective uncertainty. In: Egan A, Weatherson B (eds) Epsitemic Modality, Oxford University Press, Oxford, UK, chap 8, pp 249–269
Swanson E (2016) The application of constraint semantics to the language of subjective uncertainty. Journal of Philosophical Logic 45(2):121–146
van Benthem J (2010) Modal logic for open minds. CSLI Publications, Stanford, California
van Benthem J, Smets S (2015) Dynamic logics of belief change. In: van Ditmarsch H, Halpern JY, van der Hoek W, Kooi B (eds) Handbook of epistemic logic, College Publications, Milton Keynes, UK, chap 7, pp 313–385
von Fintel K (1999) Npi licensing, strawson entailment, and context dependency. Journal of semantics 16(2):97–148
von Fintel K, Gillies AS (2010) Must... stay... strong! Natural Language Semantics 18(4):351–383
von Fintel K, Gillies AS (2016) Still going strong: the semantics and pragmatics of epistemic *must*, manuscript for talk
Willer M (2013) Dynamics of epistemic modality. Philosophical Review 122(1):45–92
Williamson T (2000) Knowledge and its Limits. Oxford University Press, Oxford, UK
Yalcin S (2005) A puzzle about epistemic modals. MIT Working Papers in Linguistics 51:231–272
Yalcin S (2007) Epistemic modals. Mind 116(464):983–1026
Yalcin S (2010) Probabilistic operators. Philosophy Compass 5(11):916–937
Yalcin S (2011) Nonfactualism about epistemic modality. In: Egan A, Weatherson B (eds) Epistemic modality, Oxford University Press, Oxford, UK, pp 295–332