

Might do better: Flexible relativism and the QUD*

Bob Beddor
National University of Singapore

Andy Egan
Rutgers University

Submitted 2017-04-27 / First decision 2017-08-13 / Revision received 2017-10-28 /
Accepted 2018-01-26 / Final version received 2018-05-22 / Published 2018-06-05 /
Final typesetting 2024-08-20

Abstract The past decade has seen a protracted debate over the semantics of epistemic modals. According to contextualists, epistemic modals quantify over the possibilities compatible with some contextually determined group's information. Relativists often object that contextualism fails to do justice to the way we assess utterances containing epistemic modals for truth or falsity. However, recent empirical work seems to cast doubt on the relativist's claim, suggesting that ordinary speakers' judgments about epistemic modals are more closely in line with contextualism than relativism (Knobe & Yalcin 2014; Khoo 2015). This paper furthers the debate by reporting new empirical research revealing a previously overlooked dimension of speakers' truth-value judgments concerning epistemic modals. Our results show that these judgments vary systematically with the question under discussion in the conversational context in which the utterance is being assessed. We argue that this 'QUD effect' is difficult to explain if contextualism is true, but is readily explained by a suitably flexible form of relativism.

Keywords: epistemic modals, contextualism, relativism, truth-value judgments, experimental semantics

* We are grateful to participants in the 2017 Relativism and Contextualism Workshop at NYU, the 2017 meeting of the AAP, and the 2018 meeting of the Central APA for very helpful feedback. We are particularly indebted to Josh Knobe for tremendously useful discussions and for invaluable assistance designing, executing, and analyzing the experimental results. Thanks are also due to Justin Khoo, Michael Franke, and three anonymous referees at *Semantics and Pragmatics* for their detailed and thoughtful comments.

©2018 Bob Beddor and Andy Egan

This is an open-access article distributed under the terms of a Creative Commons Attribution License (<https://creativecommons.org/licenses/by/3.0/>).

1 Introduction

According to a traditional contextualist semantics, the truth-values of sentences containing epistemic modals are fixed by the context of utterance and the world of evaluation. In recent years tradition has come under fire: a number of authors have argued that it does not do justice to the conditions under which we assess such modals for truth and falsity. Many of these authors advocate replacing contextualism with a relativist semantics, according to which the truth-values of sentences containing epistemic modals also depend on a context of assessment (Egan, Hawthorne & Weatherson 2005; Egan 2007; Stephenson 2007a,b; MacFarlane 2011, 2014).

While this relativist challenge has generated much discussion, only recently have researchers begun empirically testing the predictions of contextualism and relativism. The present paper is intended as a contribution to this empirical project. Whereas previous experimental results seem to count against relativism (Knobe & Yalcin 2014; Khoo 2015), our data provide new support for the relativist camp. In particular, our studies reveal a previously overlooked dimension of speakers' judgments about the truth-values of epistemic modals: these judgments vary systematically with the question under discussion in the context of assessment. We argue that these results can be readily explained by adopting a suitably flexible form of relativism, according to which the appropriate context of assessment for evaluating an epistemic modal is constrained by assessors' conversational aims. By contrast, standard versions of contextualism have a much harder time accounting for the data.

2 Contextualism and the relativist resurgence

2.1 Contextualism and its predictions

According to a standard-issue contextualist semantics for modals, modals quantify over a set of accessible worlds, where the accessibility relation is determined by the context of utterance. Possibility modals (e.g., *might*, *possibly*) existentially quantify over such worlds; necessity modals (e.g., *must*, *necessarily*) universally quantify over them. Let ' \diamond ' stand for a possibility modal; let c be a context of utterance; and let R_c be the c -determined accessibility relation. We can formulate the standard semantics for possibility modals as follows:

Might do better

Contextualist *Might* $[\diamond\phi]^c = \{w \mid \exists w' : R_c(w, w') \& w' \in [\phi]^c\}$ ¹

When the modal is epistemic, the accessibility relation will reflect compatibility with some contextually determined body of information. On a standard interpretation, this information corresponds to the information possessed by — or available to — some contextually determined agent, or group of agents. And so an unembedded or ‘bare’ epistemic possibility claim (BEP) such as:

- (1) Simon might be in his office.

expresses, in a context of utterance c , a possible-worlds proposition that’s true at a world w just in case the prejacent (*Simon is in his office*) is compatible with the c -selected information.

While contextualism remains the default semantics for epistemic modals (at least in many quarters), one particular feature of this semantics has sparked considerable controversy. The controversial bit is this: according to Contextualist *Might*, any token utterance of a BEP will have the same truth-value for every assessor (that is, every agent evaluating the utterance for truth or falsity). And so everybody looking to make a correct assessment about the truth-value of this utterance will need to converge on the same answer. What the answer is will depend on (a) which body of information was selected by the speaker’s context, and (b) whether that information is compatible with the truth of the prejacent. As a result, assessors ought to defer to the speaker’s context; they ought to base their truth-value assessment on the compatibility of the prejacent with the information selected by the context of utterance. And so they shouldn’t evaluate the utterance for truth or falsity on the basis of whether the prejacent is compatible with their own information, except insofar as they have reason to think that their information comprises — in whole or in part — the contextually selected information state.

Given the assumption that competent assessors will (defeasibly) gravitate towards the correct judgments about the truth-values of utterances of epistemic modals, this generates the following empirical prediction:

¹ For defenses of a contextualist semantics along broadly these lines, see Kratzer 1981, 2012; DeRose 1991; Dowell 2011; Hawthorne 2007; Dorr & Hawthorne 2013; Yanovich 2014, among others.

Simple Contextualist Prediction Suppose a BEP is uttered in some particular context *c*. People assessing this claim for truth or falsity will tend to judge this claim true iff the prejacent is compatible with the *c*-selected information.

2.2 Relativism and its predictions

Over the last decade, a number of authors have contested the Simple Contextualist Prediction. To focus on one source of trouble that has loomed large in the literature, consider eavesdropper scenarios, such as the following (adapted from Egan 2007):

(2) SPYING ON SPECTRE

James Bond and Felix Leiter have just returned to London after infiltrating SPECTRE's base in the Swiss Alps. During the course of their infiltration they planted a bug, along with misleading evidence suggesting that Bond absconded to Zürich. While listening to their newly placed bug from the safety of MI6 headquarters, they overhear Blofeld and his henchman, No. 2, uncover the evidence.

- a. *No. 2*: Bond might be in Zürich.
- b. *Blofeld to No.2*: That's true.
- c. *Leiter to Bond*: That's false.

According to Egan, while it is perfectly natural for Blofeld to appraise No.2's utterance of (2a) as true, it would be much less natural for eavesdropping Leiter to do the same. Rather, Egan contends, a falsity verdict — such as (2c) — would be the far more natural reaction.²

If this contention is correct (about which more shortly), then it casts doubt on the Simple Contextualist Prediction, and, by extension, Contextual-

² For arguments in a similar vein, see Egan, Hawthorne & Weatherson 2005; MacFarlane 2011, 2014. Note that throughout this paper we will be focusing specifically on *truth-value* judgments about utterances of epistemic modals, rather than *disagreement* judgments. This difference is important: there is reason to think that our ordinary notion of 'disagreement' applies to a wide range of phenomena, from incompatibilities in assertion, to conflicts within beliefs, to clashes within preferences and desires. Absent some substantive account of disagreement, it is by no means clear that disagreement judgments — for example, the judgment that Leiter disagrees with No. 2 — count as evidence against contextualism. For further discussion of different species of disagreement and how they bear on the contextualism/relativism debate, see among others, MacFarlane 2007; Sundell 2011; Plunkett & Sundell 2013; Roberts 2017; Beddor 2018.

Might do better

ist *Might*. After all, if contextualism is true, then the difference in Leiter and Blofeld’s epistemic position vis-à-vis the proposition that Bond is in Zürich makes no difference to whether No. 2’s utterance is true. And so the difference in their inclination to appraise this utterance as true is left unexplained.

Eavesdropper arguments have been used to motivate a significant revision to the standard semantics for epistemic modals. According to relativists, a particular utterance of an epistemic modal—for example, No. 2’s pronouncement (2a)—does not have a fixed truth-value for all assessors within a world. Rather, this utterance is true for Blofeld, since its prejacent (*Bond is in Zürich*) is compatible with Blofeld’s information. But it is false for Leiter, since its prejacent is incompatible with Leiter’s information.

While there are various ways of developing a relativist semantics, one natural option is to hold that the content of a BEP (relative to a context of utterance) is a set of *centered worlds*, where a centered world is an ordered pair of a world and an assessor (alternatively, an ordered triple of a world, an assessor, and a time). Relativists can retain the idea that epistemic modals function as quantifiers over a set of accessible points; it’s just that the points are now centered worlds:

$$\text{Relativist } \mathbf{Might} \llbracket \diamond \phi \rrbracket^c = \{ \langle w, a \rangle \mid \exists \langle w', a' \rangle : R_c(\langle w, a \rangle, \langle w', a' \rangle) \& \langle w', a' \rangle \in \llbracket \phi \rrbracket^c \}$$

On this approach, an utterance of (1) (*Simon might be in his office*) expresses a centered-worlds proposition that’s true at a centered world $\langle w, a \rangle$ just in case it’s compatible with the information available to a in w that Simon is in his office.³ As a result, BEPs express a sort of content that can take different truth-values relative to different assessors within a world. Thus if Billy doesn’t know whether Simon is in his office, whereas Suzy knows that Simon is at home, $\llbracket (1) \rrbracket^c$ will be true relative to Billy, but false relative to Suzy.

How should assessors evaluate this sort of centered content for truth or falsity? Here is a natural first thought: when assessing an utterance u that expresses a set of centered worlds p , an assessor a who inhabits a world w should say *true* if $\langle w, a \rangle \in p$, and they should say *false* otherwise. So in the case of (1), assessors should base their truth-value attributions on whether *they* have access to information that rules out Simon’s being in his

³ For a relativist semantics along these lines, see Egan 2007 and Stephenson 2007a,b. For a somewhat different implementation, see MacFarlane 2011, 2014.

office, not on whether the speaker — or the speaker’s conversational group — has access to such information. Thus interpreted, relativism predicts that eavesdroppers should *not* defer to the speaker’s context when assessing a BEP for truth or falsity. Rather, they should rely on their own information.

This way of unpacking relativism delivers the following empirical prediction:

Simple Relativist Prediction Suppose a BEP is uttered in some particular context *c*. People assessing this claim for truth or falsity will tend to judge this claim true iff the prejacent is compatible with *their* information.

Taking stock: standard-issue versions of contextualism and relativism make different predictions about how ordinary people will evaluate epistemic modals for truth or falsity.⁴ Which of these predictions is correct? Relativists frequently appeal to cases like SPYING ON SPECTRE to suggest that their prediction is closer to the mark. After all, in SPYING ON SPECTRE, Leiter does not seem to defer to No. 2’s context of utterance when evaluating the truth of the BEP. Rather, Leiter appears to reject the utterance based on his own body of information — precisely what the relativist predicts. However, as we are about to see, the data are not nearly so straightforward.

3 The contextualist counterattack

3.1 Voices of discontent

Once the relativists began wheeling out their eavesdropper arguments, it did not take long for others to express misgivings. In seminars, conferences, and in print, it was not uncommon to find people who lacked the intuition

⁴ Of course, contextualism and relativism aren’t the only semantics in town. To mention one prominent alternative, a variety of authors have advanced some version of a dynamic or expressivist semantics for epistemic modals (see, among others, Veltman 1996; Beaver 2001; Gillies 2001; Swanson 2006; von Stechow & Gillies 2007; Yalcin 2007, 2011, 2015; Willer 2013; Moss 2015). For the purposes of this paper, we will for the most part set aside such approaches, since it is controversial what they predict about eavesdropper cases. (See Yalcin 2011: pp. 311–312 for the position that expressivism predicts ambivalence and unclarity in truth-value judgments when it comes to eavesdropper scenarios. See Willer 2013: pp. 60–61 for a dynamic approach that aligns more closely with the Simple Relativist Prediction.) That said, we will return to dynamic and expressivist treatments in Section 7.3, where we discuss one way of explaining our data in a dynamic framework.

Might do better

that Leiter's falsity-attribution was appropriate, or whose intuitions were murky. (See, for example, Portner 2009: p. 180; Dowell 2011: p. 14; Yalcin 2011: p. 304.) Moreover, people were quick to observe that even if some intuitions seem to favor the relativist, others seem to favor the contextualist (von Fintel & Gillies 2008, 2011; Dowell 2011; Schaffer 2011). To illustrate, consider a pair of cases that have featured prominently in the literature (Swanson 2006; von Fintel & Gillies 2008, 2011; Dowell 2011; Willer 2013):

(3) KEY SEARCH

Context: Alex and Billy are searching for their misplaced keys.

- a. *Alex:* The keys might be in the car.
- b. *Billy:* No, I actually had them when we came in the house.
- c. *Alex:* Ok, scratch that — I was wrong.

(4) KEY SQUABBLE

Context: A more acrimonious key hunt is underway.

- a. *Alex:* The keys might be in the car.
- b. *Billy:* No, I still had them when we came in the house. Why did you say that?
- c. *Alex:* Look, I didn't say that they were in the car. All I said was that they might be there.

Compared to the contextualist, the relativist has an easy time explaining KEY SEARCH — in particular, why Alex, after receiving Billy's testimony, judges her assertion of (3a) to be mistaken. After all, the prejacent (*the keys are in the car*) is inconsistent with her updated body of information; hence the modal claim is false relative to her posterior context. But unlike the contextualist, the relativist has a hard time explaining why Alex seems to be justified in "sticking to her guns" in KEY SQUABBLE. If the modal claim is false relative to her posterior context, then whence her reluctance to pronounce it false?

3.2 Experimental work to date

In light of the controversy surrounding the data, recently researchers have begun empirically investigating ordinary speakers' evaluations of utterances containing epistemic modals. In the rest of this section, we review the two main empirical investigations to date.

Knobe & Yalcin (2014) recruited participants through Amazon Mechanical Turk (AMT) and presented them with a variety of eavesdropper scenarios. While the results were rather nuanced, the overall pattern of responses suggested that participants' judgments align more closely with the Simple Contextualist Prediction than the Simple Relativist Prediction.

Their first experiment — which we will refer to as “FAT TONY 1” — used a vignette in which a mobster, Fat Tony, fakes his death in order to evade the police. Participants were told that Fat Tony places highly compelling evidence of his murder at the docks, which is then discovered by the authorities. A forensic expert is summoned to the scene. After carefully reviewing the evidence, he pronounces *Fat Tony might be dead*. Knobe and Yalcin found that participants were more inclined to agree with the statement that what the expert said is true ($M \approx 5$ on a scale from 1 ('completely disagree') to 7 ('completely agree')) than with the statement that what the expert said is false ($M \approx 2$).

As Knobe and Yalcin observe, this result stands in tension with the Simple Relativist Prediction. To see why, note that the vignette places participants in the position of better-informed assessors — assessors who know that Fat Tony is alive. And so if the Simple Relativist Prediction is right, participants' views about the truth or falsity of the expert's utterance should track whether its prejacent is compatible with their information. Since they have just read that Fat Tony is alive, they should be more inclined to agree with the statement that what the expert said is false than with the statement that what the expert said is true. But this wasn't what Knobe and Yalcin found — in fact, participants displayed the opposite inclination.

In FAT TONY 1, the vignette does not explicitly specify any assessor. In order to determine whether including an assessor in the vignette makes a difference, Knobe and Yalcin conducted a follow-up experiment, which we'll label “FAT TONY 2”. The scenario was the same as before, except that it specified that Fat Tony was watching the forensic expert on television. In one condition, participants were told that Fat Tony remarks to his henchmen, *What [the forensic expert] said is true*. In the other condition, they were told that Fat Tony remarks to his henchmen, *What [the forensic expert] said is false*. Knobe and Yalcin found that participants were more inclined to agree with Tony's claim that what the expert said is true ($M \approx 5$) than with his claim that what the expert said is false ($M \approx 3$). Here too, the results align more closely with the Simple Contextualist Prediction than with the Simple Relativist Prediction.

Might do better

Knobe and Yalcin also conducted a further experiment comparing judgments of truth and falsity to judgments about the appropriateness of *retracting* a BEP. The experiment, which we'll label "BOSTON", used an eavesdropper vignette adapted from MacFarlane (2011). In it, a speaker asserts *Joe might be in Boston* on the basis of a careful consideration of the evidence available to her. Immediately thereafter, her interlocutor receives an email saying that Joe is in Berkeley — information that he goes on to share with the speaker. Participants were more inclined to agree with the claim that it would be appropriate for the speaker to take back what she said ($M > 5$) than with the claim that what speaker said was false ($M \approx 3$).

In important follow-up work, Khoo (2015) compared judgments about truth and falsity with judgments about the appropriateness of *rejecting* a BEP. In Khoo's experiment, which we'll label "FAT TONY 3", participants were given Knobe and Yalcin's Fat Tony scenario. They were told to imagine that they are involved in the investigation as well, and that they know that Fat Tony is alive. In keeping with Knobe and Yalcin's results, Khoo found a relatively low rate of agreement with the statement that the forensic expert's claim (*Fat Tony might be dead*) is false ($M = 2.4$). However, Khoo found a comparatively high rate of agreement with the statement that it would be appropriate to reject the forensic expert's claim by saying something along the lines of: *No, Fat Tony is alive, he faked his own death* ($M = 5$). This suggests that people are more inclined to reject a BEP when the prejacent is incompatible with their information than they are to pronounce it false.

4 Relativism, revised and revived

On the face of it, these results bode ill for relativism. That said, we think it would be premature for relativists to throw in the towel. In this section, we start by raising doubts about whether the extant data really support contextualism. We then go on to advance a modified relativist hypothesis that, we contend, offers a more promising explanation of the data.

4.1 Do the data really support contextualism?

While the results reported in the previous section conflict with the Simple Relativist Prediction, at least two features of the data stand in tension with standard versions of contextualism. First, in two of the experiments (FAT TONY 2 and BOSTON) the mean rating of agreement with the claim that the

BEP is false was just over 3 on a 7-point Likert scale. This is *prima facie* surprising if contextualism is right. After all, according to contextualism, the forensic expert's utterance of *Fat Tony might be dead* expresses the proposition that Fat Tony's demise is compatible with the contextually relevant information. Presumably, Fat Tony's information is not itself included in the contextually relevant information. But then Expert B's utterance should be judged straightforwardly true, and Fat Tony's assessment of it should be judged straightforwardly false.

Second, consider the retraction and rejection data. The results of BOSTON indicate that people consider it appropriate for a speaker to retract a BEP upon learning that its prejacent is false. Similarly, the results of FAT TONY 3 indicate that people consider it appropriate for an interlocutor to reject a BEP upon learning that its prejacent is false. But neither of these predictions follow straightforwardly from standard versions of contextualism. According to contextualism, the speaker's modal utterance was true when uttered, and it remains true even after the new evidence comes in. Why, then, would it be appropriate to retract or reject it?

Taking stock: there are a number of phenomena that emerged from both the recent experiments and the pre-experimental literature that need to be explained. First, we saw in Section 3.1 that discussions in the pre-experimental literature revealed interpersonal differences in judgments regarding eavesdropper cases; they also revealed that the same authors had different judgments depending on the case (e.g., KEY SEARCH VS. KEY SQUABBLE). Second, the experimental results reviewed in Section 3.2 suggest a lower rating of agreement with the claim that BEPs are false than we would expect if the Simple Relativist Prediction were correct, but a higher rating of agreement than we would expect if the Simple Contextualist Prediction were correct. Finally, we would like to explain the differences between truth-value judgments and retraction/rejection judgments — in particular, why speakers tend to judge that it is appropriate to retract/reject an utterance of a BEP upon learning that its prejacent is false. Is there any view that can deliver the goods?

4.2 Flexible relativism

Here is a hypothesis that we find promising.⁵ Suppose we retain the basic relativist semantics (Relativist *Might*). But unlike traditional versions of rela-

⁵ We discuss alternative hypotheses in Sections 6-7.

Might do better

tivism we deny that assessors invariably rely on their own context of assessment when assessing the truth-value of a BEP. Instead, they can look to other contexts of assessment — for example, that of the speaker. Call the resulting view “flexible relativism.”

Here is a way to think about this sort of view. The proposition that’s expressed by an utterance of *Might p* is a centered worlds proposition — the one corresponding to the property: *lacking information that rules out p*. Asked to deliver a verdict about the truth or falsity of the utterance, there are different questions one could be seeking to answer. One could be offering an answer to the question, *is the proposition expressed true of me (the assessor)?* Or one could be offering an answer to the question, *is the proposition expressed true of the speaker?* In the first case — where the better-informed assessor is delivering a verdict about whether the proposition expressed is true of themselves — they should say *false* if they know that the prejacent is false. In the second case — where the better-informed assessor is delivering a verdict about whether the proposition expressed is true of the speaker — they should say *true* as long as the prejacent is compatible with the speaker’s information. Possibly there are other questions they could be addressing as well, which would lead their truth-value attributions to track (their beliefs about) the truth-value of the centered worlds proposition at other contexts of assessment.

Flexible relativism is relativism shorn of the Simple Relativist Prediction. It is thus consistent with all of the experimental data discussed in Section 3.2. As it stands, however, flexible relativism is hardly a satisfying theory. While flexible relativism is consistent with the observed data, it does not yet generate any predictions about when people will use their own context of assessment and when they will look elsewhere. This gives rise to the worry that it is *too* flexible; it purchases compatibility with the data at the expense of predictive power.⁶

In order to convert flexible relativism into a predictive theory, we propose supplementing it with pragmatic constraints imposed by the conversational context. Crucially, however, the relevant conversational context is *not* the context in which the epistemic modal was initially uttered, but rather the *context of assessment*.

⁶ MacFarlane (2011: Section 8.2) briefly discusses a version of flexible relativism. Khoo (2015: p. 531) also mentions the possibility of adopting flexible relativism in passing. However, to our knowledge no one has developed flexible relativism into a genuinely predictive theory. One of the aims of the present paper is to fill this lacuna.

This brings us to a further observation about the studies summarized in the previous section. While all of the vignettes provided some conversational context for the utterance of the BEP, none of them provided a context for participants to use when assessing the utterance for truth or falsity. Of course, one of the experiments (FAT TONY 2) provided a context in which a character pronounces on the truth-value of the earlier utterance. But in being asked to state whether they agree or disagree with this pronouncement, participants are being asked to assess an assessment. And no conversational context was provided for making this further assessment.

How does conversational context determine which context of assessment to use when evaluating a BEP? It is likely that a number of different features of the conversational context will influence the choice of a context of assessment. However, we hypothesize that one of the primary constraints is provided by the question under discussion (QUD):

QUD Constraint Suppose someone in a conversational context c is assessing an utterance of a BEP for truth or falsity. *Ceteris paribus*, they will be inclined to assess the utterance using whichever context of assessment is most relevant to answering the QUD in c .⁷

The ‘*ceteris paribus*’ clause is important here: it would be implausibly strong to claim that the QUD Constraint is the *only* principle governing the choice of a context of assessment. Another plausible candidate is a *charity*

⁷ The idea that conversations are guided by a stack of questions under discussion that the interlocutors are mutually committed to resolving has proven influential in explaining a wide range of linguistic phenomena, from prosody to presupposition projection to ambiguity and anaphora resolution. For an overview of the question under discussion framework, see Roberts 1996–2012.

In recent work, Roberts (2023, 2017) discusses the impact of the QUD on the interpretation of epistemic modals. In many respects Roberts’ discussion complements our own. However, there are important differences. One is that Roberts’ account is contextualist: for her, the extension of a BEP is only sensitive to the QUD in the context of utterance. For reasons that will become clear in Sections 5 and 6, we think that the experimental results reported here motivate a relativist view, according to which the extension of a BEP is sensitive to the QUD in the context of assessment. Another difference is that while Roberts focuses on disagreement, we focus on ascriptions of truth and falsity. This is not merely a terminological difference: as Roberts (2017: p. 20) notes, one can be in a position to register disagreement with someone without being in a position to claim that they spoke falsely. And so even if contextualists can provide a fully satisfactory explanation of disagreement judgments, it’s not clear that they can leverage this into an explanation of truth-value judgments.

Might do better

constraint, according to which listeners will be inclined to use whichever context of assessment will make the utterance of the BEP come out true. Sometimes the various constraints governing the choice of a context of assessment will align, joining their voices in unanimous support of some particular context of assessment. But sometimes these constraints will pull apart. When this happens, we may expect assessors to be ambivalent; we may also expect interpersonal variation in how assessors prioritize these constraints.

Our formulation of the QUD Constraint raises at least two questions. First, what happens when the context of assessment does not contain a clear QUD, or there are multiple equally relevant contexts of assessment for answering the QUD? In this case, we should expect some degree of ambivalence and variability. Ambivalence, because those uncertain about the QUD are liable to be uncertain about which response to offer. Variation, because in the absence of sufficient cues to determine a QUD, different assessors may make different guesses about the QUD, leading to variation in the context of assessment selected. That said, there may well be a tendency to default to the speaker's context in the absence of further contextual cues. After all, in scenarios where the QUD Constraint provides no clear guidance, it would be natural for assessors to assign greater weight to other constraints, such as a charity constraint. In standard eavesdropper cases, a charity constraint recommends deferring to the speaker's context, since this will make the utterance of the BEP come out true.

Second, what explains the fact that assessments of truth or falsity are sensitive to the QUD in this way? Does this sensitivity derive from some more general pragmatic principle, or from the semantics of truth and falsity attributions, or both? Officially, flexible relativism is compatible with different ways of answering this question. However, one promising option is to explain the QUD-sensitivity of truth attributions by positing a hidden argument place in the truth predicate. According to this proposal, claims of the form, *p is true* are elliptical for claims of the form, *p is true at context of assessment c* (and similarly for falsity claims). One way of formally implementing this suggestion would be to hold that the truth predicate takes two arguments, a centered worlds proposition *p*, and a centered world,⁸ delivering the value *true* just in case the centered world belongs to *p*. That is:

Relativist True $\llbracket is\ true \rrbracket^{c,\langle w,a \rangle} = \lambda p.\lambda w'.\lambda x.p(w')(x) = 1$

⁸ Alternatively: a function from centered worlds to centered worlds.

On this approach, the QUD influences how we fill in this hidden argument place — in particular, whether we fill it in with the assessor’s context of assessment or the speaker’s. This is in keeping with a growing body of research indicating that the QUD plays an important role in resolving ambiguity and ellipsis.⁹ As such, the QUD Constraint is not a pragmatic rule specifically tailored to epistemic modals; rather, it follows from a much more general constraint governing how listeners assign meanings to utterances.

Given these clarifications, we are now in a position to see how flexible relativism offers a predictive theory. It predicts that variations in the QUD in the context of assessment will result in variations in judgments about the truth-values of BEPs. In particular, suppose that an assessor A’s information can be used to answer the QUD in A’s conversational context, but the speaker’s information cannot. Then we should expect — all else being equal — A to be inclined to use A’s context rather than the speaker’s. And so we should expect A to evaluate an utterance of a BEP in line with the Simple Relativist Prediction: A will be inclined to judge it true iff the prejacent is consistent with A’s information. In standard eavesdropper cases, this will happen when the QUD in A’s context concerns the truth or falsity of the prejacent. After all, in standard eavesdropper cases A is better-informed than the speaker regarding the truth-value of the prejacent: A knows that the prejacent is false, whereas the speaker does not.¹⁰

By contrast, suppose the speaker’s information is more relevant to answering the QUD in A’s conversational context. Then we should expect — all else being equal — A to be inclined to defer to the speaker’s context. And so we should expect A to evaluate an utterance of a BEP in line with the Simple

⁹ See e.g., Gualmini et al. 2008 and Zondervan, Meroni & Gualmini 2008 on the role of the QUD in resolving ambiguities; see Clifton & Frazier 2012, Grant, Clifton & Frazier 2012, and Elliot, Nicolae & Sudo 2014 on the role of QUD in resolving ellipsis. For discussion of how the QUD impacts the interpretation of context-sensitive expressions, see Roberts 2003 and Schaffer & Szabó 2014.

¹⁰ This is certainly the structure of SPYING ON SPECTRE, FAT TONY, KEY SEARCH/SQUABBLE, and many of the other cases discussed in literature. Of course, we can also imagine ‘ignorant assessors’ scenarios (Dietz 2008) in which the assessor does not know whether the prejacent is true, but knows that someone else does. In such a scenario, the QUD Constraint predicts that assessors will be inclined to defer to the better-informed agent. This prediction seems to be correct. We can imagine Watson saying, *If Holmes said, ‘There’s no possibility that the gardener did it’, then what he said is true.* Here Watson is evaluating Holmes’ hypothetical utterance relative to Holmes’ context, since Holmes’ information is more relevant to answering the question, *Did the gardener do it?* See MacFarlane 2011: Section 8.2 for further discussion of how flexible relativism handles such cases.

Might do better

Contextualist Prediction: A will be inclined to judge it true iff the prejacent is consistent with the speaker's information. This will often happen when the QUD in A's conversational context is a normative question about the speaker — for example, whether the speaker's initial inquiry was competent, or whether the speaker's assertion of the modal claim was warranted.¹¹

4.3 Explaining the data to date

In the next section, we report the results of a series of new experiments that bear out these predictions. But before presenting new results, it is worth considering how the QUD Constraint enables flexible relativists to explain the data collected to date. According to the explanation that emerges, differences in people's intuitions about the truth or falsity of BEPs are attributable, at least in part, to differences in how people interpret the QUD in the context of assessment. Those whose judgments about a particular case align with the Simple Relativist Prediction are often interpreting the QUD as concerning the truth or falsity of the prejacent. Those whose judgments about a particular case align with the Simple Contextualist Prediction are often interpreting the QUD differently — for example, as concerning the competence of the speaker. The commonly-reported ambivalence and murkiness in people's intuitions is attributable, at least in part, to the fact that many cases in the literature are under-described. In particular, many of the cases fail to provide enough cues for readers to clearly identify a QUD in the context of assessment.¹²

¹¹ This isn't the only kind of situation where speaker's context is most relevant to addressing the QUD. We might also be directly interested in discovering which propositions are consistent with the speaker's information. Such questions are particularly liable to arise when the speaker's information is complicated and it is easy to lose track of which propositions are consistent with it. Consider, for example, von Stechow and Gillies' 'Mastermind' scenario (2007: 45, 2008: 83) in which the guesser says, *There might be two reds*, and the better-informed player replies, *That's true*, meaning that the prejacent is consistent with the guesser's information. (Thanks to an anonymous referee for pointing out the connection with this sort of case.)

¹² Dowell (2011) helpfully draws attention to the intuitional murkiness and ambivalence concerning the cases used by relativists to challenge contextualism about epistemic modals. Dowell also makes the point, which we here endorse, that one thing that's gone wrong in many such cases, and is responsible for much of the squishiness and variability of people's responses to them, is that they are underspecified in important ways. We diverge from Dowell in our diagnosis of which bits of additional specification of the cases are going to be crucial in fixing responses. Dowell emphasizes fixing details about (a) context of utterance,

At the same time, flexible relativists can also explain why there is a greater tendency in under-described cases to judge the speaker's utterance true than to judge it false. As noted above, it is plausible that in the absence of a clear QUD, assessors will be guided by a charity constraint. And this will recommend deferring to the speaker's context in eavesdropper scenarios.¹³

Flexible relativism also explains why different cases seem to elicit different judgments — in particular, why some eavesdropper cases tend to elicit relativist intuitions whereas others tend to elicit contextualist intuitions.¹⁴ Consider again the contrast between KEY SEARCH and KEY SQUABBLE. In KEY SEARCH, the conversation is focused on the location of the keys. Hence the QUD is whether the prejacent of Alex's assertion of (3a) (*The keys might be in the car*) is true. Presumably, this remains so even after Billy asserts (3b) (*No, I actually had [the keys] when we came in the house*). Given this, the QUD Constraint explains why it is appropriate for Alex to pronounce her initial assertion false after hearing Billy's assertion: the most relevant context of assessment for answering the QUD is her current context of assessment ($Alex_{posterior}$), which incorporates Billy's testimony, and hence rules out the prejacent. By contrast, in KEY SQUABBLE Billy does not merely inform Alex: *I still had them when we came in the house*. Billy also adds a petulant *Why did you say that?* In doing so, Billy implicitly raises the question of whether Alex was warranted in her initial assertion. And the context of assessment that is most relevant to answering *this* question is Alex's previous context of assessment ($Alex_{prior}$), which did not rule out the prejacent.¹⁵

and (b) assessors' beliefs about the context of utterance. We emphasize the fixing details about the context of assessment.

- 13 Recall that participants' judgments were more mixed when an eavesdropper scenario included an explicit assessment of a modal utterance (FAT TONY 2) than when it did not (FAT TONY 1). One explanation for this difference is that charity considerations unambiguously favor deferring to the forensic expert's context in FAT TONY 1 (since the expert is the only speaker), whereas in FAT TONY 2 participants are forced to choose between being charitable towards the first speaker (the forensic expert) or the second speaker (Fat Tony).
- 14 This variation in judgments across cases was not directly demonstrated by any of the experiments summarized in Section 3.2. However, discussions of eavesdropper cases in the non-experimental literature points to the existence of such variation (Section 3.1).
- 15 This explanation extends to handle some of the other cases that have been offered against relativism. Take Dowell's (2011: pp. 16–17) lottery scenario, in which A asks B, *Why did you buy that ticket?*, to which B replies, *I might win*. The next day the winner is announced, and an empty-handed B finds himself confronted by A, who demands, *Why did you buy that ticket yesterday? You didn't win!* As Dowell observes, it would be odd for B to concede, *You're right. What I said yesterday was false*. The QUD Constraint offers an explanation. When A

Might do better

Flexible relativism also offers a way to explain judgments about the appropriateness of rejection and retraction, and why these come apart from judgments about truth or falsity. Consider: what is it to accept some proposition p ? Here's a natural thought: for someone to accept p is for them to take p to be true relative to their current context. What is it to reject p ? It is to decline to accept it. And so if A knows that the content of B's utterance is false relative to A's current context, then A will reject the content of B's utterance. Moreover, it will typically be appropriate for A to publicly signal this rejection by using a negation marker such as *No*.

Finally: what is it to retract some earlier claim? It is to publicly signal that one no longer accepts it, or no longer advocates others accepting it. And so if someone discovers that their earlier claim is false relative to their current context, then it should be appropriate for them to retract it. (After all, they no longer accept its content, and presumably they no longer wish to advocate for their current interlocutors to accept its content.) On the picture that emerges, judgments about truth-value are flexible, in that they can track truth/falsity in different contexts of assessment. By contrast, judgments about the normative status of accepting/rejecting/retracting a claim are inflexible, in that they invariably track truth in the context of assessment of the person doing the accepting/rejecting/retracting.¹⁶

5 Epistemic modals and the QUD: Experimental results

In this section we report on two experimental studies testing whether truth-value assessments are QUD-sensitive in the way that flexible relativism predicts. In both studies, all participants were given a vignette in which a speaker asserts *Might p* on the grounds that p is consistent with the information at hand. At a later time, an assessor who knows that p is false is asked whether

challenges B with the question, *Why did you buy that ticket yesterday?*, A raises the question of whether B's ticket-buying was reasonable. And the context of assessment that is most relevant to answering *this* question is the context that B inhabited prior to the lottery draw. Hence the oddity of B's concession.

¹⁶ This isn't the only possible way of thinking about acceptance, rejection, and retraction within a relativist framework. One could instead propose that accepting p is a matter of taking p to be true relative to the most *relevant* context (or contexts), where what counts as 'relevant' is determined by the QUD—likewise with rejection and retraction. (See [Spencer 2016](#) for a relativist view of belief that is closer to this approach.) We think the results of BOSTON and FAT TONY 3 count against this alternative construal. For further discussion of what relativists should say about rejection and retraction, see [MacFarlane 2011, 2014](#): ch. 5.

the speaker’s assertion is true. Each participant was randomly assigned to one of two conditions that differ only with respect to the QUD in the context of assessment. In the QUD-PREJACENT condition, it is made clear that the point of the truth-value query is to determine whether p is itself true; the interest in determining the truth-value of speaker’s assertion is entirely subordinated to this goal. In the QUD-COMPETENCE condition, it is made clear that the point of the truth-value query is to determine whether the speaker conducted their initial investigation competently; the interest in determining the truth-value of the speaker’s assertion is entirely subordinated to this goal. In both conditions, participants were asked which would be the correct response — saying that the utterance of the BEP is true, or saying that the utterance is not true. We predicted that we would see a difference between the two conditions: participants in the QUD-PREJACENT condition would be more inclined than participants in the QUD-COMPETENCE condition to deny that the utterance is true.

5.1 Experiment 1: John’s medical visit

5.1.1 Methods

In the first experiment (devised in cooperation with, and conducted on AMT by, Josh Knobe), 104 participants responded to a vignette that began as follows:

John had a rare disease, so he went to the doctor to ask about medications.

The doctor said, “Researchers have tested a whole bunch of different medications for this disease, but if you look at all of the published studies, you can see that every single medication they have tried hasn’t worked at all.” Then the doctor started talking about a new medication called accuphine. He said, “Accuphine might cure you. There is a new study being run right now to test it out, but they haven’t released the results yet, so there is no way to know whether it works or not.”

Participants received one of two continuations of this vignette, depending on which condition they had been assigned. Those who had been assigned to the QUD-PREJACENT condition received a continuation containing a conversation centered on whether accuphine will cure John:

Might do better

QUD-PREJACENT CONDITION

John carefully looked at all of the publically available research, and sure enough, judging just by the publically available research, there was no way to know whether accuphine worked or not. But now John wanted to know whether he should take accuphine. John happened to be friends with one of the medical researchers who was working on the study. So he decided to go talk to the medical researcher.

The medical researcher knew something that John didn't. She knew that the study had been completed and that it had been determined that accuphine didn't work at all. In fact, it had been discovered that if John took it, he would probably end up being seriously harmed.

John told the medical researcher that his doctor had said "Accuphine might cure you." Then John asked the medical researcher, "Is what my doctor said true?"

Participants assigned to the QUD-COMPETENCE condition received a continuation containing a conversation centered on the doctor's competence:

QUD-COMPETENCE CONDITION

John happened to be friends with one of the medical researchers who was working on the study. She told him that the study had been completed and that it had been determined that accuphine didn't work at all. In fact, it had been discovered that if John took it, he would probably end up being seriously harmed. But now John wanted to know whether his doctor was basically good at his job or whether he should maybe switch over to a different doctor. So he decided to go talk to the medical researcher again.

The medical researcher knew something that John didn't. She had carefully looked at all of the publically available research, and sure enough, if one judged just by the publically available research, there was no way to know whether accuphine worked or not.

John told the medical researcher that his doctor had said "Accuphine might cure you." Then John asked the medical researcher, "Is what my doctor said true?"

In both conditions, participants were given the following forced choice question:

The medical researcher can now answer either “True” or “False.” We want to know which of these two answers you think is the correct one. So please tell us which answer is correct.

___ True ___ False

5.1.2 Experiment 1: Results and discussion

The results for each condition are displayed in Figure 1. A higher percentage of the participants picked *False* in the QUD-PREJACENT condition (85%) than in the QUD-COMPETENCE condition (67%). The difference between the two conditions was significant, $\chi^2(1, N = 104) = 4.67, p = .031$.¹⁷

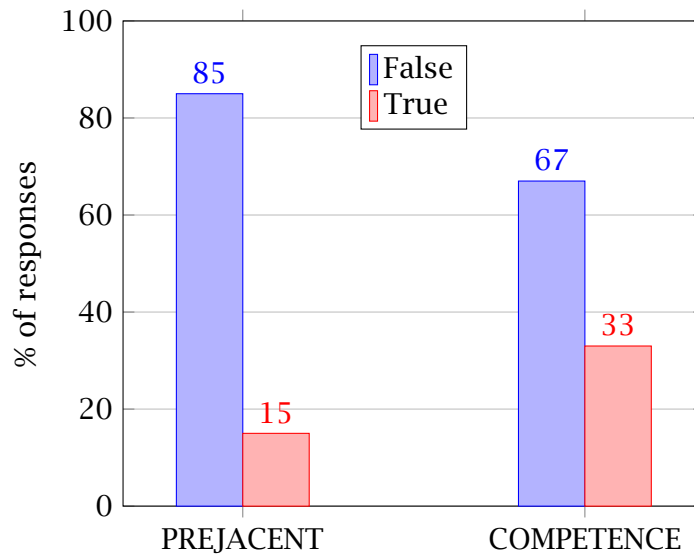


Figure 1 Proportion of responses by condition in Experiment 1

Two aspects of these results are worth highlighting. First, they suggest that assessments of the truth-values of utterances containing epistemic

¹⁷ The complete data for all of the experiments reported here can be found at <https://github.com/rbeddor/QUDepestemicmodals>.

Might do better

modals are sensitive to the QUD in the context of assessment — as predicted by flexible relativism supplemented with the QUD Constraint.

Second, the results are much more in line with the Simple Relativist Prediction than the results of previous experimental studies. Particularly surprising is the fact that even in the QUD-COMPETENCE condition (which was hypothesized to be more in line with the Simple Contextualist Prediction), the majority of the participants still selected *False*. Why is this? At least two possibilities strike us as plausible. One is that participants assigned to the QUD-COMPETENCE condition failed to register the QUD, or lost track of the QUD by the time they reached the end of the vignette. Another possibility is that the high stakes nature of the scenario makes it difficult to shift the QUD away from the question of whether the prejacent is true. (Recall that participants are told that if John takes accuphine he would probably be seriously seriously harmed.)¹⁸

5.2 Experiment 2: Fat Tony's return

In order to investigate whether the high proportion of participants who selected *False* was attributable to some anomalous features of the medical vignette — such as the high stakes, or the difficulty in tracking the QUD — we decided to replicate the same basic experimental design using a different scenario. For our second experiment, we developed a variant of the FAT TONY scenario employed by Knobe, Yalcin, and Khoo. In developing our variant, we strove to make the QUD in both conditions as explicit as possible.

5.2.1 Methods

125 participants were recruited through AMT. Each received a vignette that began as follows:

¹⁸ An anonymous reviewer raises a third possibility, which is that the future tense of the modal (*Accuphine might cure you*) influenced the results. In particular, perhaps future tense *might*-claims are naturally read as expressing some sort of objective or metaphysical possibility, rather than epistemic possibility (cf. Werner 2006). A full assessment of how these objective/metaphysical readings of *might*-claims influence truth-value judgments will depend on the semantics of such objective/metaphysical modals. However, we agree that this is a potential confound, one that motivates using present tense modals in further experiments.

Fat Tony is a mobster who has faked his own death in order to evade the police. He secretly plants highly compelling evidence of his murder at the docks. The evidence is discovered by the authorities, and the forensic expert is summoned to the scene. After carefully reviewing the evidence he concludes, "Fat Tony might be dead."

Participants received one of two continuations of this vignette, depending on which condition they were assigned. Those assigned to the QUD-PREJACENT condition received a continuation of the vignette containing a discussion centered around whether Fat Tony is dead:

QUD-PREJACENT CONDITION

A week later, a murder is committed. The detective assigned to investigate the murder is trying to figure out whether Fat Tony could be responsible and consults a criminal informant. The informant knows that Tony is alive.

Detective: "We're trying to find the killer. We know that the victim was an enemy of Fat Tony. However, our forensic expert said, 'Fat Tony might be dead.' Is what he said true?"

Participants assigned to the QUD-COMPETENCE condition received a continuation of the vignette containing a discussion of whether the forensic expert (named "Ed" in this version of the vignette) conducted his investigation competently:

QUD-COMPETENCE CONDITION

A week later, another forensic expert, Ted, is given new evidence that conclusively shows that Fat Tony is still alive. The police department is trying to determine whether the initial investigation was competent, and so sends a detective to interview Ted.

Detective: "We're trying to figure out whether Ed's initial investigation was competent. On the basis of the initial evidence, Ed said, 'Fat Tony might be dead.' Is what he said true?"

In both conditions, participants were given the following forced choice question:

Might do better

Which of the following responses would be correct?

- (a) “No, it’s not.”
- (b) “Yes, it is.”

5.2.2 Results and discussion

Five out of 125 participants were excluded for failing the comprehension check.¹⁹ Analyses were conducted on the remaining 120 responses. A higher percentage of participants (59%) selected (a) (*No, it’s not*) in the QUD-PREJACENT condition than in the QUD-COMPETENCE condition (33%). (See [Figure 2](#).) The difference between the conditions was found to be highly significant, $\chi^2(1, N = 120) = 8.23, p = .004$.

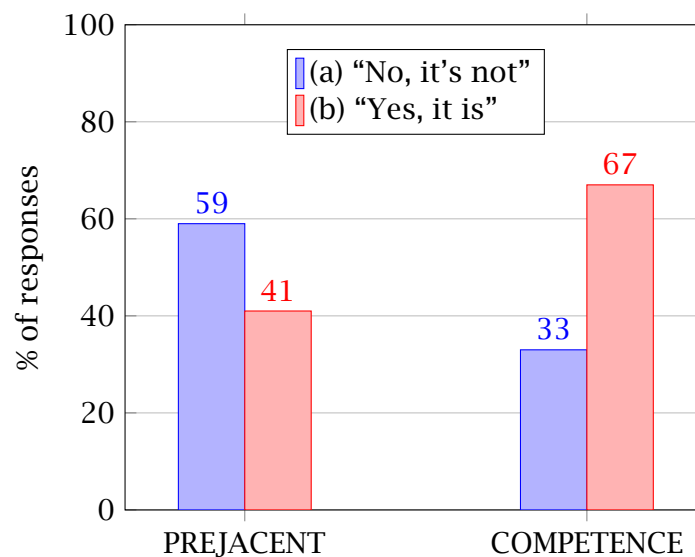


Figure 2 Proportion of responses by condition in Experiment 2

These results provide further evidence of a QUD effect: truth-value judgments about BEPs are sensitive to the QUD of the conversational context in which the utterance is being assessed. Whereas in Experiment 1 the majority of participants in both conditions judged the target utterance false, in Experiment 2 varying the QUD changed the majority judgment. (The majority

¹⁹ For the comprehension check, we required subjects to answer the question: *In the scenario you were given, was Fat Tony alive?* (The same comprehension check was also used for Experiments 3 and 4.)

of participants in the QUD-PREJACENT condition judged the target utterance false; the majority of participants in the QUD-COMPETENCE condition judged it true.) This variation is explained by flexible relativism, supplemented with the QUD Constraint.

One further feature of the results is worth discussing. While the majority of participants in both conditions selected the option predicted by the QUD Constraint, a very substantial minority selected the other option (41% in the QUD-PREJACENT condition, 33% in the QUD-COMPETENCE condition). What explains this? As in the case of Experiment 1, this may be due, in part, to difficulty in identifying the QUD. While we attempted to make the QUD as explicit as possible, the vignettes are (unavoidably) rather complicated, which may interfere with readers' ability to keep the QUD in mind. However, another hypothesis is also worth exploring — one that does not require attributing any sort of error to the participants. According to this alternative hypothesis, some participants are relying on additional constraints governing the choice of the appropriate context of assessment, constraints that they prioritize over the QUD Constraint. For example, the 41% of participants who selected (b) in the QUD-PREJACENT condition may be relying on a charity constraint (Section 4), which leads them to select the context of assessment that will make the speaker's utterance come out true.

In order to test which of these hypotheses is correct, we conducted a follow-up experiment testing participants' ability to track the QUD in the context of assessment. A follow-up experiment along these lines is independently desirable, since it would help corroborate one of the central claims of flexible relativism. In particular, flexible relativism, when supplemented with the QUD Constraint, predicts that assessors are sensitive to the QUD in the context of assessment, but we have not yet directly tested this prediction.²⁰

5.3 Tracking the QUD

5.3.1 Experiment 3: Methods

In order to test whether participants were able to track the QUD, we ran a variant on Experiment 2. As before, we recruited participants through AMT and randomly assigned them to either a QUD-PREJACENT or a QUD-COMPETENCE condition. The vignettes were exactly the same as in Experiment 2; the only

²⁰ Thanks to an anonymous referee for drawing this concern to our attention.

Might do better

difference was the question that we asked. Rather than asking which response would be correct, we instead asked:

In this scenario, which question is the detective most interested in finding out the answer to:

- (a) Whether Fat Tony is alive.
- (b) Whether the initial forensic investigation at the docks was competent.

5.3.2 Experiment 3: Results and discussion

The responses of five of 122 participants were discarded for failing the comprehension check; we analyzed the remaining 117 responses. 89% of participants assigned to QUD-PREJACENT condition selected (a) (*Whether Fat Tony is alive*), whereas only 20% of participants assigned to QUD-COMPETENCE condition selected (a). (See Figure 3.) The difference between the conditions was found to be extremely significant, $\chi^2(1, N = 117) = 55.78, p < .0001$.

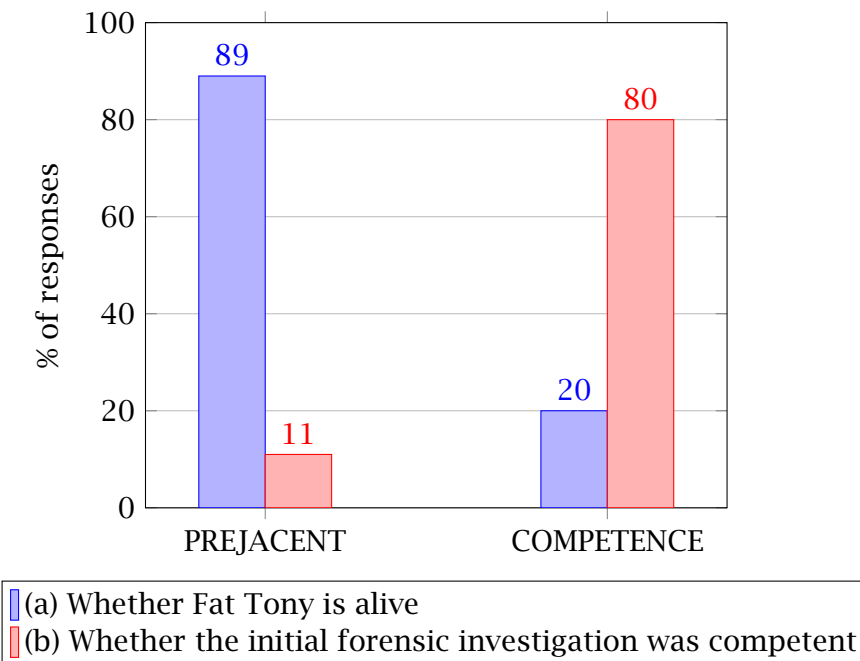


Figure 3 Proportion of responses by condition in Experiment 3

Two aspects of these results are worth emphasizing. First, they provide strong evidence that participants are able to reliably track the QUD in the context of assessment, at least when the QUD is made fairly explicit. And so they alleviate a potential concern for our explanation of the QUD effect, which is that we lack independent reason to think that participants are sensitive to the QUD in the first place.

Second, we see a stronger consensus in QUD judgments (within each condition) than in truth-value judgments (as reported in Experiment 2). Of course, we do see *some* within-condition variation in participants' judgments about the QUD. And so these results are consistent with the hypothesis that divergences in QUD judgments partially explain within-condition divergences in truth-value judgments. But the comparatively strong consensus in QUD judgments suggests that this isn't the full story.

This provides further reason to think that the QUD Constraint is not the sole constraint guiding the choice of a context of assessment. Other constraints — for example, a charity constraint — will often come into play, and people may differ in how they prioritize these constraints. Fully developing this proposal would require giving a more detailed account of the additional constraints governing the choice of the context of assessment than we can offer here. However, it strikes us as plausible that listeners will differ along these lines, and that this follows from a more general phenomenon — namely, that people differ in the strategies they use for resolving ambiguities.²¹

6 Can contextualists explain the QUD effect?

In the previous section, we've offered experimental results suggesting that speakers' judgments about the truth-values of utterances containing epistemic modals are subject to a QUD effect: they vary systematically with the QUD in the context of assessment. We've also argued that this effect is readily explained by flexible relativism. But is there any way of explaining this effect within a contextualist framework? In this section, we review three candidate contextualist explanations and argue that none of them succeeds. While this

²¹ Cf. [Gualmini et al. 2008](#), which explains differences in the way children and adults interpret scopally ambiguous sentences in terms of differences in how they prioritize two constraints on utterance interpretation: a charity constraint and the Question-Answer Requirement (which requires that listeners interpret utterances as answering the QUD).

Might do better

does not prove that no contextualist explanation of the data is forthcoming, it does suggest that developing such an explanation is by no means trivial.²²

6.1 The Prejacent-Targeting Hypothesis

One way contextualists could try to explain our data is to suggest that variations in the QUD do not affect the choice of a context of assessment. Rather, they affect the appropriateness of *targeting the prejacent*.

Let's unpack this idea. Sometimes when we respond to an assertion with a negation marker, what we intend to deny is not the assertion itself, but rather the assertion's prejacent (Simons 2007). For example:

- (5) a. *Alex*: I heard that Sarah will come to the party.
- b. *Billy*: No!/Nuh uh!/That's not true!

On one available interpretation, Billy is not denying that Alex *heard* that Sarah will come to the party. Rather, Billy is denying that Sarah will come to the party.

As various authors have observed, this phenomenon of “prejacent-targeting” can explain some of the data that relativists take to be problematic for contextualists (von Stechow & Gillies 2008: pp. 82–83, 2011: p. 115; Dowell 2011: p. 14). Recall the exchange from KEY SEARCH:

- (3) a. *Alex*: The keys might be in the car.
- b. *Billy*: No, I actually had them when we came in the house.

Here too it's plausible that Billy is denying the prejacent of Alex's assertion (*The keys are in the car*) rather than the assertion itself.

Whether a negation marker is most naturally interpreted as targeting the prejacent of an assertion or the assertion itself depends on the QUD. Contrast two contexts for (5): in the first, the conversation is centered on who will attend the party; in the second, the conversation is centered on the various things that Alex has been told throughout the day. In the first context, Billy's utterance of (5b) is most naturally interpreted as denying the prejacent of (5a). In the second, Billy's utterance is most naturally interpreted as denying the entire assertion (that Alex heard that Sarah will come to the party).

²² Here we focus on whether your garden-variety contextualism (of the sort discussed in Section 2.1) can account for the data. In the next section, we discuss a non-standard “cloudy contextualist” treatment.

What goes for embedding verbs such as *heard* also goes for modals: there too, varying the QUD will influence the acceptability of targeting the prejacent. And this, contextualists may suggest, yields a natural explanation of the results of Experiments 1 and 2. When the QUD is whether the prejacent is true, it will be appropriate to target the prejacent; hence assessors will be inclined to judge an utterance of a BEP to be false when they know the prejacent is false. When the QUD concerns the speaker's competence, it will be less appropriate to target the prejacent; hence assessors will be more inclined to judge an utterance of a BEP to be true provided the prejacent is consistent with the speaker's information.

On the face of it, this "Prejacent-Targeting Hypothesis" would appear to provide an elegant contextualist-friendly explanation of the results of our studies. Moreover, this hypothesis seems independently plausible, since even in non-modal constructions such as (5) the QUD influences whether we interpret a negation marker as targeting the prejacent of an assertion or the assertion itself.

6.2 How to test the Prejacent-Targeting Hypothesis

Let us start by raising a qualm about the Prejacent-Targeting Hypothesis. We fully agree that in many contexts a negation marker can be used to target an assertion's prejacent. But this observation does not suffice to explain the data in the previous section. To explain the data, one must also maintain that a negation marker can be used to target an assertion's prejacent *even when the negation marker is uttered in response to a question about the truth-value of the assertion*. And this seems far from obvious. Consider:

- (6) *Context*: Some friends are musing about who will attend tonight's soiree.
- a. *Charlie*: Alex said that she heard that Sarah will come to the party. Is what she said true?
 - b. *Billy*: No, it's not.

Is there an available interpretation on which Billy is denying that Sarah will come to the party? This is far from clear to us. But this is precisely what the Prejacent-Targeting Hypothesis claims is going on in our studies.

That said, intuitions about such examples are admittedly a bit murky, and may not provide a particularly firm basis for adjudicating between flexible

Might do better

relativism and the Prejacent-Targeting Hypothesis. It is thus worth exploring whether we can test the Prejacent-Targeting Hypothesis directly.

One strategy would be to construct a case in which an assessor A is in a position to know that a modal claim is false (relative to A's context of assessment), but not in a position to know that the prejacent itself is false. If in the QUD-PREJACENT condition A is still inclined to judge the modal utterance false, this is evidence against the Prejacent-Targeting Hypothesis. After all, by construction of the case, A is not in a position to deny the prejacent. Hence the Prejacent-Targeting Hypothesis would afford no explanation of A's judgment.

Can we construct such a case? Not with *might*, or at least not obviously — after all, given Relativist *Might*, an utterance of *Might p* is false relative to A's context of assessment only if A's information is inconsistent with *p*. But if A's information is inconsistent with *p*, then A should be in a position to know that *p* is false. And so we will never have a case where *Might p* is false relative to A's context of assessment, but A does not know whether *p* is false.

However, we can construct such a case if we switch from *might* to *probably*. For contextualists, an utterance of *Probably p* is true iff *p* has a sufficiently high probability conditional on some body of information selected by the context of utterance. For relativists, *Probably p* is true relative to a context of assessment iff *p* has a sufficiently high probability conditional on the assessor's information (or, more generally, the information of some group determined by the context of assessment).²³

By switching to *probably*, we can easily construct a case of the desired form. All we need is to build a scenario where (i) *p* has a high probability conditional on the speaker's information, (ii) there is some assessor A who is better-informed than the speaker regarding *p*, (iii) *p* has a low (but non-zero) probability conditional on A's information. In such a case, flexible relativism about *probably* predicts that A will judge the speaker's utterance of *Probably*

²³ A simple formal implementation: let f_c be a contextually determined function from a world (or a centered world) to a set of accessible worlds (or centered worlds). Let t be some threshold between 0 and 1 (perhaps also determined by c). Then we can formulate the contextualist and relativist semantics for *probably* as follows:

Contextualist *Probably* $\llbracket \text{Probably } \phi \rrbracket^c = \{w \mid \text{Pr}(\llbracket \phi \rrbracket^c \mid f_c(w)) > t\}$

Relativist *Probably* $\llbracket \text{Probably } \phi \rrbracket^c = \{\langle w, a \rangle \mid \text{Pr}(\llbracket \phi \rrbracket^c \mid f_c(\langle w, a \rangle)) > t\}$

For relevant discussion of the semantics of probability operators, see Swanson 2006; Lassiter 2010, 2017; Yalcin 2010; Klecha 2012; Moss 2015.

p to be false in the QUD-PREJACENT condition. (After all, A's context will be the most relevant context for determining whether p is true, since A is better-informed than the speaker. Given a relativist semantics, *Probably* p will be false relative to A's context, since p is stipulated to have a low probability conditional on A's information.) By contrast, the Prejacent-Targeting Hypothesis does not predict that the assessor will judge the speaker's utterance of *Probably* p to be false in the QUD-PREJACENT condition.

We now report the results of two further experiments that bear out the flexible relativist hypothesis on this point.

6.3 Testing the Prejacent-Targeting Hypothesis: Experiment 4

6.3.1 Methods

For our first test of the Prejacent-Targeting Hypothesis we developed another variant of the FAT TONY scenario. 240 participants were recruited through AMT and randomly assigned to one of two conditions: QUD-PREJACENT and QUD-COMPETENCE. The vignettes they received were the same as in Experiment 2, except for two important differences. First, the original forensic expert (Ed) says, *Fat Tony is probably dead* after reviewing the evidence at the docks, rather than *Fat Tony might be dead*. Second, the assessor (the criminal informant in the QUD-PREJACENT condition, the second forensic expert (Ted) in the QUD-COMPETENCE condition) does not know for sure that Fat Tony is alive; all they know is that it is likely that he is alive. (The criminal informant knows this from talking to his criminal accomplices; Ted knows this on the basis of new evidence that has come to light.) As before, participants in both conditions were given the detective's question (*Is what he [Ed] said true?*) and asked to select which response would be correct: (a) *No, it's not* or (b) *Yes, it is*.

6.3.2 Results and discussion

The responses of 21 out of 245 participants were discarded for failing the comprehension check. We ran the analysis on the remaining 224 responses. 69% of participants assigned to the QUD-PREJACENT condition selected (a) (*No, it's not*), whereas 53% of participants assigned to the QUD-COMPETENCE condition selected (a). (See [Figure 4](#).) The difference between the two conditions was found to be significant, $\chi^2(1, N = 224) = 6.06, p = .014$.

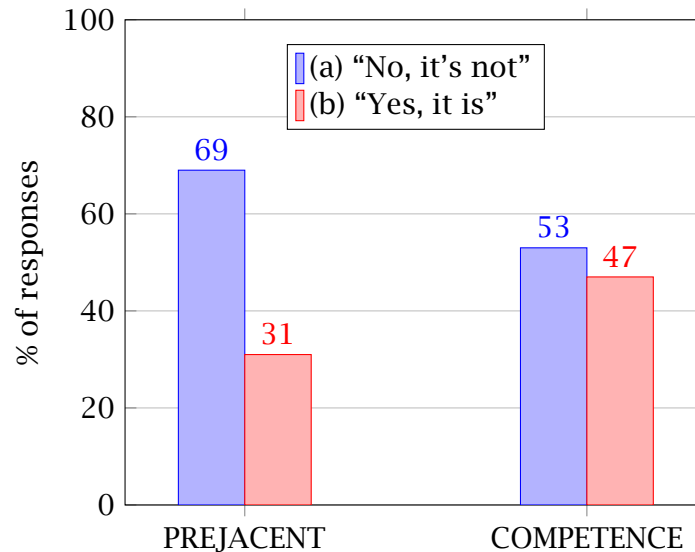


Figure 4 Proportion of responses by condition in Experiment 4

Two observations about these results are worth highlighting. First, these results extend the main finding of the previous section. Not only are truth-value judgments about BEPS subject to a QUD effect; the same goes for *probably* claims. Second, and more importantly for our purposes, these results provide evidence against the Prejacent-Targeting Hypothesis. After all, the Prejacent-Targeting Hypothesis predicts that participants will not be inclined to select (a) in either condition, since in both conditions the assessor does not know that the prejacent is false. However, the majority of participants were inclined to select (a) — indeed, they were inclined to select (a) in both conditions (about which more momentarily).

By contrast, flexible relativism readily explains why the majority of participants in the QUD-PREJACENT condition selected (a). After all, the most relevant context for answering the question, *Is Fat Tony alive?* is the criminal informant’s context, since the criminal informant has information that the expert — Ed — lacks. Given a relativist semantics for *probably*, Ed’s utterance of *Fat Tony is probably dead* is false relative to the CI’s context, since the prejacent has a low probability conditional on the CI’s information.

However, the results in the QUD-COMPETENCE condition are more surprising for flexible relativism. Why did slightly more than half of the participants in this condition judge the modal utterance false? Part of this could be due to factors already discussed: for example, perhaps they are relying on addi-

tional constraints when selecting a context of assessment for evaluating the modal utterance. But another factor may also be affecting the results in this case: *probably* is arguably a vague expression, and participants may disagree about how much evidence for p is required to warrant a claim of *Probably p*. Of course, participants were told that the evidence at the docks is “highly compelling”; however, the expression *highly compelling* is also vague, and so perhaps some participants are interpreting it in such a way that highly compelling evidence for p does not necessarily entail that probably p . In order to address this potential confound, we ran a follow-up experiment using explicit numerical probabilities.

6.4 Testing the Prejacent-Targeting Hypothesis: Experiment 5

6.4.1 Methods

To make the assignment of explicit probabilities as natural as possible, our second experiment employed a medical scenario in which John receives an initial test for strep throat, and later receives a follow-up test (a throat culture). 249 participants were recruited through AMT and received a vignette that began thus:

John is worried he might have strep throat. He goes to his primary care physician and she runs an initial test that indicates that there is a 75% chance that John does not have strep. Based on the initial test results, John’s doctor says: “You probably don’t have strep throat. However, we should do a throat culture in order to be safe. If it turns out that you have strep throat, we should put you on antibiotics.”

Participants were randomly assigned to either a QUD-PREJACENT condition or a QUD-COMPETENCE condition, and received a different elaboration of the case depending on which condition they were assigned. Those assigned to the QUD-PREJACENT condition received a continuation in which the conversation centers on whether John has strep throat:

QUD-PREJACENT CONDITION

John comes back two days later to find out the results of the throat culture, and sees a different doctor. The throat culture comes up positive, which indicates there is a 90% chance that

Might do better

John has strep throat. John has not yet seen the results of these tests, but his new doctor has. John asks the new doctor: “I’m trying to figure out whether I need to take antibiotics. My primary care physician told me, ‘You probably don’t have strep.’ Is what she said true?”

Which of the following responses would be correct?

- (a) “No, it’s not”
- (b) “Yes, it is”

Those assigned to the QUD-COMPETENCE condition received a continuation in which the conversation centers on the competence of John’s primary care physician:

QUD-COMPETENCE CONDITION

John comes back two days later to find out the results of the throat culture, and sees a different doctor. The throat culture comes up positive, which indicates there is a 90% chance that John has strep throat. But now John wants to know whether his primary care physician made a mistake administering the initial test, so he asks: “I’m trying to figure out whether I can rely on my primary care physician. She told me, ‘You probably don’t have strep’. Is what she said true?”

The new doctor reviews the initial tests, and confirms that John’s primary care physician had not made any mistakes interpreting the results. Given this, which of the following responses would be correct?

- (a) “No, it’s not”
- (b) “Yes, it is”

6.4.2 Results and discussion

The responses of 28 of 249 participants were discarded for failing the comprehension check.²⁴ We analyzed the remaining 221 responses. 74% of the participants assigned to the QUD-PREJACENT condition selected (a) (*No, it’s*

²⁴ For the comprehension check, we required participants to answer to the question, *In the scenario you just read, which of the following did John’s primary care physician say?* The

not) whereas only 39% of those assigned to the QUD-COMPETENCE condition selected (a). (See Figure 5.) The difference between the two conditions was found to be extremely significant, $\chi^2(1, N = 221) = 27.47, p < .0001$.

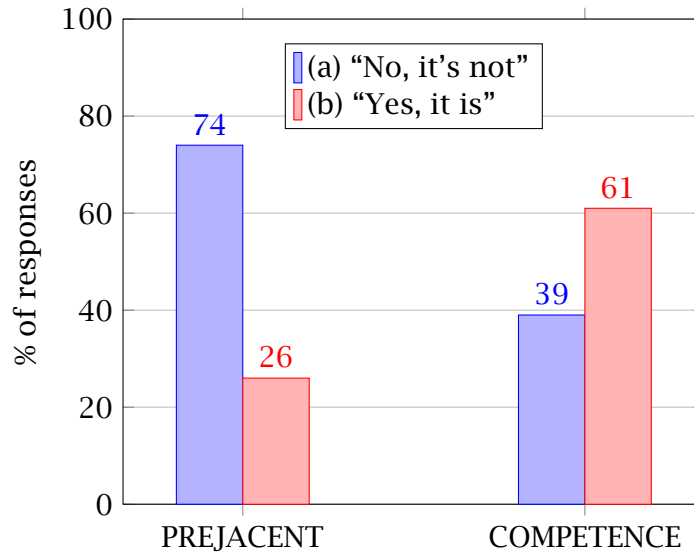


Figure 5 Proportion of responses by condition in Experiment 5

These results provide further evidence that truth-value judgments about *probably* claims are subject to a QUD effect.²⁵ They also provide further evidence against the Prejacent-Targeting Hypothesis. If the Prejacent-Targeting Hypothesis were correct, participants in both conditions should be disinclined to select (a). After all, the new doctor does not know for sure whether the prejacent (*John does not have strep throat*) is false; they only know that there is a 90% chance that it is false. However, a very substantial majority of participants selected (a) in the QUD-PREJACENT condition.²⁶

options were: (a) *You probably have strep.* (b) *You probably don't have strep.* (c) *You probably have pneumonia.* (d) *You probably don't have pneumonia.* (The same comprehension check was used for Experiment 6, discussed in Section 6.6.)

²⁵ Note also that in this case, in which explicit numerical probabilities are provided, the majority of respondents in the QUD-COMPETENCE condition selected (b).

²⁶ One other feature of this experiment merits comment. In this probe — as in our other experiments — the truth-elicitation question is present-tensed (*Is what she said true?*) rather than past-tensed (*Was what she said true?*). There is an interesting question as to whether changing the tense would affect the results. One reason for using the present tense is that it is not clear whether flexible relativism makes different predictions from contextualism when

Might do better

6.5 The Wrong Question Hypothesis

We take the results of Experiments 4 and 5 to cast serious doubt on the Prejacent-Targeting Hypothesis. Is there any other contextualist explanation of the QUD effect?

An alternative hypothesis is that some of our participants are simply answering the wrong question. To motivate this hypothesis, contextualist may observe that the vignettes we've used are fairly complex. In particular, each vignette contains two questions. The first is the QUD; the second is about the truth-value of a modal utterance. While participants are only asked to answer the second question, it is not implausible that some may be confused, and answer the first question instead. Indeed, such a mistake would be particularly likely — and particularly understandable — if we succeeded in our aim of making the QUD as salient as possible; the more salient the QUD, the more likely that this is the question participants will latch onto. If participants are confused in this way, it would explain why the majority select (a) (*No, it's not*) in the QUD-PREJACENT condition; these participants would be interpreting the anaphor *it* as referring to the prejacent rather than the modal assertion. Call this the “Wrong Question Hypothesis.”

While this is a natural concern, there are at least two reasons for being skeptical of this explanation of the data. The first is that this explanation is an error theory: it claims that many — perhaps most — participants in the QUD-PREJACENT condition are making a mistake. It may be an understandable mistake, but it is a mistake nonetheless. Now while it is certainly possible that speakers are systematically mistaken in this way, we think it is a theoretical cost to posit such rampant speaker error. In saying this, we merely mean to endorse a commonly assumed criterion of charity, according to which theories that make speakers' linguistic judgments come out true should be preferred, *ceteris paribus*, to those that do not. By this criterion, flexible relativism should be preferred to the Wrong Question Hypothesis.

A second reason for doubting the Wrong Question Hypothesis is that it does not explain the results of Experiments 4 and 5. Recall that these vignettes were set up in such a way that the assessor in the QUD-PREJACENT condition (the criminal informant in Experiment 4, the second doctor in Experiment 5) does not know whether the prejacent is true. Hence in these vignettes, the assessor is not in a position to answer the QUD in the QUD-

it comes to past tense truth-value judgments. After all, flexible relativists could maintain that tense mandatorily shifts the relevant context of assessment.

PREJACENT condition. However, the majority of participants still selected (a) in both experiments.

These reservations notwithstanding, it is worth trying to test the Wrong Question Hypothesis directly. Fortunately there is a simple way to do so: simply modify the possible answers. Instead of asking participants to choose between the options *No, it's not* and *Yes, it is*, we could ask them to choose between options that explicitly make reference to the speaker's assertion:

Modified Question Which of the following responses would be correct?

- (a) "No, what she said isn't true"
- (b) "Yes, what she said is true"

Here there is no possibility of misinterpreting either option as a pronouncement on the truth of the prejacent. In what follows, we report the results of a follow-up study relying on this modified answer choice.

6.6 Testing the Wrong Question Hypothesis: Experiment 6

6.6.1 Methods

For simplicity, we re-ran one of our previous experiments (Experiment 5); the only change was that participants were given the Modified Question above. (While we used Experiment 5 as our basis, we could equally well have chosen any of the other experiments.) As before, participants were recruited through AMT and randomly assigned to either the QUD-PREJACENT condition or the QUD-COMPETENCE condition.

6.6.2 Results and discussion

The responses of 28 out of 257 participants were discarded for failing the comprehension check; we analyzed the remaining 229 responses. 73% of the participants assigned to the QUD-PREJACENT condition selected (a) (*No, what she said isn't true*), whereas only 17% of participants assigned to the QUD-COMPETENCE condition selected (a). (See [Figure 6](#).) The difference between the two conditions was extremely significant, $\chi^2(1, N = 229) = 69.67, p < .0001$.

These results are similar to those obtained in Experiment 5 using the original question. Particularly important for our purposes are the results in

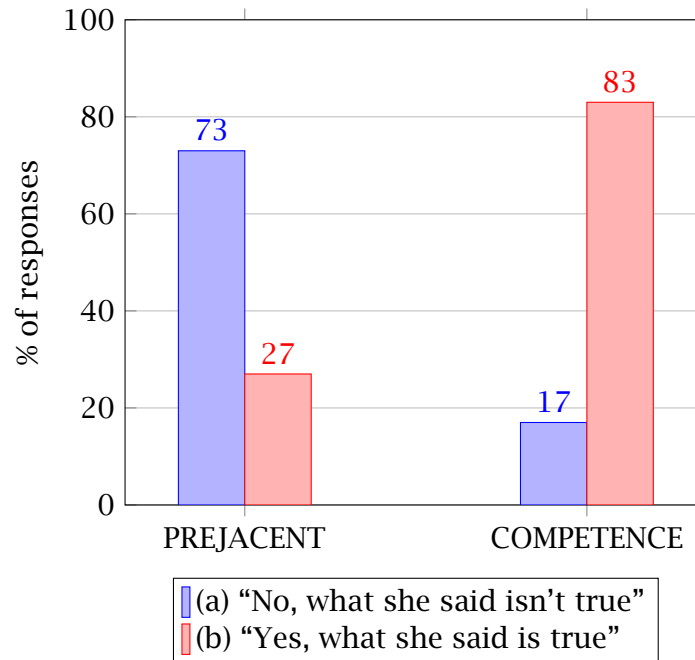


Figure 6 Proportion of responses by condition in Experiment 6

QUD-PREJACENT condition: the overwhelming majority judged that what the speaker said is not true, which runs contrary to the Wrong Question Hypothesis.²⁷

6.7 The Content Interpretation Hypothesis

Let us now consider one final contextualist explanation of the data. Contextualists typically allow that epistemic modals are flexible, in the sense that a context of utterance can select any number of different bodies of information (Dowell 2011; Yanovich 2014; Roberts 2023, 2017). Some contexts are ‘solipsistic’: they only select the speaker’s information state. Others are more inclusive: they select the collective information of all conversational participants, or the information that is in principle available to conversational participants, etc. And so in order for an assessor to accurately assess a BEP for truth or falsity, they will need to make a conjecture about which information

²⁷ Note that this result also provides further evidence against the Prejacent-Targeting Hypothesis. After all, if A asks, *Is what B said true?* and C replies, *No, what B is said is not true*, then C is unambiguously denying B’s assertion rather than its prejacent.

state was selected by the context of utterance. Perhaps, some contextualists may suggest, features of the context of assessment — in particular, the QUD — influence assessors' conjectures about which information state was selected by the context of utterance.

One way of developing this suggestion is to convert the QUD Constraint into a constraint on the assignment of contents to modal utterances:

QUD Constraint on Content Suppose someone occupying a conversational context c is trying to determine which proposition was expressed by an utterance of an epistemic modal in some context c' . *Ceteris paribus*, they will be inclined to select the interpretation of the content of the utterance that is most relevant to answering the QUD in c .

This proposal is consistent with all of our data. But is it plausible? Let us walk through how this would work in the strep throat scenario. Presumably, the thought is that when John's primary care physician says *You probably don't have strep throat*, there are at least two propositions she could be trying to convey. One is that the probability of the prejacent (*You don't have strep throat*) is high, conditional on the results of the initial test — call this proposition, 'INITIAL'. The other is that the probability of the prejacent is high, conditional on the results of both the initial test and whatever follow-up tests are performed — call this proposition, 'BOTH'. The QUD Constraint on Content predicts that when John wants to know whether he has strep, the second doctor will be inclined to interpret the primary care physician's utterance as having expressed BOTH, since this is the most relevant proposition for answering John's question. But is BOTH really a plausible candidate for the content of the primary care physician's utterance? At the time of utterance, the primary care physician did not know the results of the follow-up. And so at the time of utterance, she was not in a position to assert BOTH, and presumably would recognize as much. (If asked, *Do you mean that it's probable that I don't have strep, given the follow-up test?*, she presumably would answer, *Of course not, we don't know the results yet.*) Moreover, it seems that the second doctor should be in a position to recognize all of this, and hence to rule out BOTH as a plausible candidate for the content of the primary care physician's utterance.

We can frame the worry in more general terms. In eavesdropper cases, an assessor often has information that was unavailable to the speaker. In such cases, the speaker would not have been in a position to assert that p is compatible with (or probabilified by) the information in question. Yet according

Might do better

to the QUD Constraint on Content, assessors will often interpret speakers as having made such claims. And so if the QUD Constraint on Content were correct, we would often interpret speakers as expressing propositions that both they (the speakers) and we (the assessors) realize they were not in a position to assert.²⁸

7 Beyond flexible relativism: Alternative implementations

7.1 Taking stock

When we examine ordinary speakers' responses to eavesdropping cases, we don't see the clear and unequivocal assessment on the basis of assessors' information that early relativist arguments presupposed. But we also don't see the clear and unequivocal assessment on the basis of speakers' information (or speaker-relevant groups' information) that simple contextualist views predict. What we see instead is variability. But this variability is not without rhyme or reason. The results discussed here show that assessments of the truth-values of utterances containing epistemic modals vary systematically with QUD in the conversation in which the assessment takes place. And this phenomenon is readily explained by a flexible relativist view.

Of course, many contextualists also hold that the QUD influences the extensions of utterances containing epistemic modals (Dowell 2011; Yanovich 2014, Roberts 2023, 2017). Crucially, however, these flexible contextualists will delegate all influence to the QUD in the *context of utterance*: the QUD in the context of utterance will determine the truth-value of the utterance at a world w , and this truth-value will be invariant across all assessors that inhabit w . The difference between flexible contextualism and flexible relativism emerges clearly in cases where we hold the context of utterance fixed, but vary the QUD in the context of assessment. Flexible relativism predicts that such variations will affect people's judgments about the truth-value of

²⁸ Is there any way for contextualists to avoid this consequence? One option would be to hold that our truth-value assessments of epistemic modals do not depend on *our* concerns (in the context of assessment), but rather on our best guess about which proposition that speaker was trying to express (cf. Dowell 2013). We think this is a natural position for contextualists to adopt, and that it sidesteps the problems facing the QUD Constraint on Content. But this position won't explain the experimental results reported here. As we have seen, the most plausible candidate for the proposition that the primary care physician was trying to express is INITIAL, not BOTH. And so this position would not explain people's tendency to judge the primary care physician's utterance false in the QUD-PREJACENT condition.

the utterance of the epistemic modal. Contextualism, by contrast, predicts no such effect.

The experiments summarized in Sections 5-6 confirm the flexible relativist's predictions. While contextualists might try to explain the results of Experiments 1 and 2 by appealing to the idea that participants are evaluating the truth of the prejacent — either because of prejacent-targeting, or because they are answering the wrong question — this suggestion does not account for the results in Experiments 4-6. After all, in Experiments 4-6 the vignettes were constructed to ensure that the assessor does not know whether the prejacent is true. Moreover, in Experiment 6 the answers explicitly make reference to *what the speaker said*, rather than relying on the potentially ambiguous anaphor *it*. Nonetheless, we still observed a statistically significant QUD effect in each experiment.

While we take our results to count against standard contextualism and to count in favor of flexible relativism, we do not go so far as to claim that flexible relativism is the only framework that explains these results. Indeed, we wish to conclude by sketching two alternative frameworks — a “cloudy contextualist” framework and a dynamic framework — that may also be able to explain the QUD effect.

7.2 QUD-sensitive cloudy contextualism

“Cloudy contextualism” (developed by von Stechow & Gillies (2011)) agrees with standard-issue contextualism about the meaning of *might*: an utterance of a BEP expresses, relative to a context of utterance c , a proposition that's true at a world w iff the prejacent is compatible with the c -selected body of information. However, cloudy contextualists add a twist: often the facts on the ground do not determine a unique context, and hence do not determine a unique body of information. When this happens, there will be a number of different admissible contexts, and hence a number of different admissible information states. When the context is underspecified in this way, an utterance u of an epistemic modal does not assert a single proposition. Rather, it “puts in play” a “cloud” of different propositions, specifically, the propositions expressed by u in each of the admissible contexts.

Let us model a “cloudy context” as a set of admissible contexts $c_1 \dots c_n$ — the set of contexts compatible with the facts on the ground. We can then formulate the cloudy contextualist treatment of *might* as follows:

Might do better

$$\text{Cloudy Might } \llbracket \diamond \phi \rrbracket^{c_1 \dots c_n} = \left\{ \begin{array}{l} \{w \mid \exists w' : R_{c_1}(w, w') \& w' \in \llbracket \phi \rrbracket^{c_1}\} \\ \dots \\ \{w \mid \exists w' : R_{c_n}(w, w') \& w' \in \llbracket \phi \rrbracket^{c_n}\} \end{array} \right\}$$

On this account, an utterance of (1) (*Simon might be in his office*) in a cloudy context might put in play a solipsistic reading of the modal—according to which the prejacent (*Simon is in his office*) is compatible with the speaker’s information. But it might simultaneously put in play various more expansive group readings—for example, a reading according to which the prejacent is compatible with the information available to either the speaker or her interlocutors.

This is not, however, the full story. In order to make predictions about truth-value appraisals, cloudy contextualists supplement their semantics with various pragmatic norms. For example, [von Stechow & Gillies \(2011\)](#) offer a norm that goes roughly like this:

Cloudy Appraisal Norm Suppose someone utters a BEP, thereby putting in play propositions $p_1 \dots p_n$. Then a hearer H can appraise the utterance as true (false) if the strongest p_i that H reasonably has an opinion about is such that H thinks it is true (false).

Developed thus, cloudy contextualism does not explain the QUD effect. After all, the Cloudy Appraisal Norm makes no reference to the QUD in the conversation in which the modal assertion is being appraised for truth or falsity. As a result, it does not predict that the QUD will influence hearers’ truth-value judgments about the modal assertion.

However, an alternative norm may fare better. Suppose we replace the Cloudy Appraisal Norm with a norm that makes explicit reference to the QUD, for example:

QUD-Sensitive Appraisal Norm Suppose someone utters a BEP, thereby putting in play propositions $p_1 \dots p_n$. Then a hearer H who occupies a context (cloudy or otherwise) c should appraise the utterance as true (false) if the p_i that is most relevant to answering the QUD in c is such that H thinks it is true (false).

Equipped with this replacement, cloudy contextualism is in a good position to predict the experimental data offered here. Take Experiment 2. When the forensic expert at the docks—Ed—says, *Fat Tony might be dead*, he

puts in play a cloud of propositions. This cloud includes a solipsistic reading (*Tony's passing is compatible with Ed's information*). But it also includes various group readings — for example, an informant-inclusive group reading (*Tony's passing is compatible with both Ed and the CI's information*). In the QUD-PREJACENT condition, the QUD of the detective and the informant's discussion is whether Fat Tony is alive. And so the QUD-Sensitive Appraisal Norm predicts that the informant should pronounce the original utterance false, since the member of the cloud that is most relevant to answering this question is the informant-inclusive group reading. By contrast, in the QUD-COMPETENCE condition, the QUD of the detective and Ted's conversation is whether Ed's initial investigation was competent. And so the QUD-Sensitive Appraisal Norm predicts that Ted should pronounce the original utterance true, since the proposition that is most relevant to answering this question is the solipsistic reading. Similar remarks apply, *mutatis mutandis*, to the other experiments.²⁹

7.3 QUD-sensitive Update Semantics

One might also try to explain the QUD effect in a dynamic or expressivist setting.³⁰ As a proof of concept, we will focus on one standard dynamic semantics for epistemic modals, Update Semantics (Veltman 1996), though the basic strategy could be readily adapted to other dynamic or expressivist frameworks.

According to Update Semantics, the meaning of a sentence is identified with its context change potential (CCP) — its ability to impact the conversational context. Formally, a context c is taken to be a set of worlds, and a

²⁹ The QUD-Sensitive Appraisal Norm may have other advantages over the Cloudy Appraisal Norm. MacFarlane (2010) objects that cloudy contextualists seem to invoke *sui generis* pragmatic rules governing the appropriateness of asserting, confirming, and denying utterances involving epistemic modals — rules that do not seem to derive from any more general pragmatic principle. Arguably, the QUD-Sensitive Appraisal Norm improves on the Cloudy Appraisal Norm in this regard. After all, the QUD-Sensitive Appraisal Norm follows from a more general principle governing how to appraise utterances with underspecified or ambiguous content:

Generalized QUD-Sensitive Appraisal Norm Suppose someone makes an utterance u , thereby putting in play propositions $p_1 \dots p_n$. Then a hearer H who occupies a context (cloudy or otherwise) c should appraise u as true (false) if the p_i that is most relevant to answering the QUD in c is such that H thinks it is true (false).

³⁰ See fn. 4 for references.

Might do better

CCP is a function from contexts to contexts. The interpretation of a language is then given by an update function $[\cdot]$ from sentences to CCPs. An atomic sentence α shrinks the context to contain only worlds where α is true:

Atomic Update $c[\alpha] = \{w \in c \mid w(\alpha) = 1\}$

Whereas an atomic sentence updates the context with the worlds in which it is true, modal sentences perform tests on the context. In particular, an utterance of a BEP tests whether the context is compatible with the prejacent. If so, the context is left unchanged. If not, the context crashes, yielding the empty set:

Dynamic Might $c[\diamond\phi] = \begin{cases} c & \text{if } c[\phi] \neq \emptyset \\ \emptyset & \text{otherwise} \end{cases}$

For example, an utterance of (1) (*Simon might be in his office*) in a context c will either return c (if c contains at least one world where Simon is in his office), or the empty set (if it doesn't).

Taken by itself, *Dynamic Might* does not tell us how assessors will appraise BEPs for truth or falsity. But one could develop such a semantics in a way that mimics the predictions of flexible relativism. Here's how. Suppose a BEP is uttered in a context c . And suppose some assessor A is evaluating this utterance in a context c' . We could propose that A will judge the original utterance true iff the most relevant context passes the test imposed by the modal. As before, the most relevant context will be determined by the QUD in c' (rather than c). When the QUD is whether the prejacent is true, the most relevant context will typically be A's conversational context (c'). In this case, A will evaluate the utterance as true iff c' is compatible with the prejacent. When the QUD concerns the speaker's competence, the most relevant context will typically be the original context of utterance (c). In this case, A will evaluate the utterance as true iff c is compatible with the prejacent. This way of developing a dynamic approach yields the same predictions about truth-value judgments as flexible relativism, and hence explains the data in much the same way.³¹

³¹ As in the case of flexible relativism, one way of implementing this approach would be to hold that truth attributions have a hidden argument place. In the present framework the hidden argument place will be for a context, here taken to be a set of worlds. On this view, a claim of the form, *ψ is true* will be elliptical for a claim of the form, *ψ is true at context c* . We could then combine this with an off-the-shelf dynamic definition of truth in terms of

7.4 Concluding remarks

We take the QUD effect uncovered here to pose a problem for traditional versions of contextualism, and to motivate flexible relativism. However, we do not claim that the data decisively favor flexible relativism; as we have seen, one can also develop QUD-sensitive versions of cloudy contextualism and Update Semantics that handle the data equally well. But note that they do so only by virtue of being empirically indistinguishable from flexible relativism, at least when it comes to predictions about speakers' assessments of the truth and falsity of utterances of epistemic modals. The choice between these frameworks will thus have to be settled on other grounds — for example, other data concerning epistemic modals,³² or supra-empirical considerations such as theoretical parsimony.

References

- Beaver, David. 2001. *Presupposition and assertion in dynamic semantics*. Stanford: CSLI Publications. <https://semanticsarchive.net/Archive/jU1MDVmZ/>.
- Beddor, Bob. 2018. Subjective disagreement. *Noûs* 53(4). 819–851. <https://doi.org/10.1111/nous.12240>.
- Beddor, Bob & Simon Goldstein. 2018. Believing epistemic contradictions. *Review of Symbolic Logic* 11(1). 87–114. <https://doi.org/10.1017/s1755020316000514>.

support (see, e.g., von Stechow & Gillies 2007: 50), where a context c supports ψ iff updating c with ψ has no effect on c (i.e., $c[\psi] = c$). This gives the following semantics for truth ascriptions:

Dynamic True $c[\psi$ is true at c'] = $\{w \in c \mid c'[\psi] = c'\}$

Putting all of this together: a claim of the form, *Might ϕ is true* is elliptical for a claim of the form, *Might ϕ is true at c'* . And this claim shrinks the context to only contain worlds where c' supports *Might ϕ* , and hence to only include worlds where c' is compatible with ϕ . (That is, $c[\diamond\phi$ is true at $c'] = \{w \in c \mid c'[\phi] \neq \emptyset\}$.)

³² For example, a natural next step is to look beyond bare epistemic modals to embedded epistemic modals. For a corpus-based investigation of embedded epistemic modals, see Hacquard & Wellwood 2012. For discussion of how the behavior of embedded epistemic modals bears on the choice of semantic framework, see, among others, Yalcin 2007, 2015; Dorr & Hawthorne 2013; Willer 2013; Beddor & Goldstein 2018; Ninan 2018.

Might do better

- Clifton, Charles & Lyn Frazier. 2012. Discourse integration guided by the 'Question Under Discussion'. *Cognitive Psychology* 65(2). 352-379. <https://doi.org/10.1016/j.cogpsych.2012.04.001>.
- DeRose, Keith. 1991. Epistemic possibilities. *Philosophical Review* 100(4). 581-605. <https://doi.org/10.2307/2185175>.
- Dietz, Richard. 2008. Epistemic modals and correct disagreement. In García Carpintero & Kölbel (eds.), *Relative truth*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199234950.003.0011>.
- Dorr, Cian & John Hawthorne. 2013. Embedding epistemic modals. *Mind* 122(488). 867-913. <https://doi.org/10.1093/mind/fzt091>.
- Dowell, J.L. 2011. A flexible contextualist account of epistemic modals. *Philosophers' Imprint* 11(14). 1-25. <https://doi.org/2027/spo.3521354.0011.014>.
- Dowell, J.L. 2013. Flexible contextualism about deontic modals: A puzzle about information-sensitivity. *Inquiry* 56(2-3). 149-178. <https://doi.org/10.1080/0020174X.2013.784464>.
- Egan, Andy. 2007. Epistemic modals, relativism, and assertion. *Philosophical Studies* 133(1). 1-22. <https://doi.org/10.1007/s11098-006-9003-x>.
- Egan, Andy, John Hawthorne & Brian Weatherson. 2005. Epistemic modals in context. In Gerhard Preyer & Georg Peter (eds.), *Contextualism in philosophy*, 131-170. Oxford: Oxford University Press.
- Elliot, Patrick, Andreea Nicolae & Yasutada Sudo. 2014. The sticky reading: VP ellipsis without parallel binding. *Semantics and Linguistic Theory (SALT)* 24. 640-655. <https://doi.org/10.3765/salt.v24i0.3639>.
- von Fintel, Kai & Anthony Gillies. 2007. An opinionated guide to epistemic modality. In Tamar Gendler & John Hawthorne (eds.), *Oxford studies in epistemology*, chap. 2, 32-62. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780199237067.003.0002>.
- von Fintel, Kai & Anthony Gillies. 2008. CIA leaks. *Philosophical Review* 117(1). 77-98. <https://doi.org/10.1215/00318108-2007-025>.
- von Fintel, Kai & Anthony Gillies. 2011. 'Might' made right. In Andy Egan & Brian Weatherson (eds.), *Epistemic modality*, chap. 3, 108-130. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199591596.003.0004>.
- Gillies, Anthony. 2001. A new solution to Moore's paradox. *Philosophical Studies* 105(3). 237-250. <https://doi.org/10.1023/A:1010361708803>.

- Grant, Margaret, Charles Clifton & Lyn Frazier. 2012. The role of non-actuality implicatures in processing elided constituents. *Journal of Memory and Language* 66(1). 326–343. <https://doi.org/10.1016/j.jml.2011.09.003>.
- Gualmini, Andrea, Sarah Hulsey, Valentine Hacquard & Danny Fox. 2008. The question-answer requirement for scope assignment. *Natural Language Semantics* 16(3). 205–237. <https://doi.org/10.1007/s11050-008-9029-z>.
- Hacquard, Valentine & Alexis Wellwood. 2012. Embedding epistemic modals in English: A corpus-based study. *Semantics and Pragmatics* 5(4). 1–29. <https://doi.org/10.3765/sp.5.4>.
- Hawthorne, John. 2007. Eavesdroppers and epistemic modals. *Philosophical Issues* 17(1). 92–101. <https://doi.org/10.1111/j.1533-6077.2007.00124.x>.
- Khoo, Justin. 2015. Modal disagreements. *Inquiry* 58(5). 511–534. <https://doi.org/10.1080/0020174X.2015.1033005>.
- Klecha, Peter. 2012. Positive and conditional semantics for gradable modals. In Ana Aguilar Guevara, Anna Chernilovskaya & Rick Nouwen (eds.), *Proceedings of Sinn und Bedeutung 16*, vol. 2, 363–376. Massachusetts Institute of Technology Working Papers in Linguistics (MITWPL). <https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/433>.
- Knobe, Joshua & Seth Yalcin. 2014. Epistemic modals and context: Experimental data. *Semantics and Pragmatics* 7(10). 1–21. <https://doi.org/10.3765/sp.7.10>.
- Kratzer, Angelika. 1981. The notional category of modality. In Hans-Jürgen Eikmeyer & Hannes Rieser (eds.), *Words, worlds, and contexts: New approaches in word semantics*, 38–74. Berlin: W. de Gruyter. <https://doi.org/10.1515/9783110842524-004>.
- Kratzer, Angelika. 2012. *Modals and conditionals*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199234684.001.0001>.
- Lassiter, Daniel. 2010. Gradable epistemic modals, probability, and scale structure. *Semantics and Linguistic Theory (SALT)* 20. 197–215. <https://doi.org/10.3765/salt.v20i0.2557>.
- Lassiter, Daniel. 2017. *Graded modality*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198701347.001.0001>.
- MacFarlane, John. 2007. Relativism and disagreement. *Philosophical Studies* 132. 17–31. <https://doi.org/10.1007/s11098-006-9049-9>.
- MacFarlane, John. 2010. *Epistemic modals: Relativism vs. cloudy contextualism*. Chambers Philosophy Conference on Epistemic Modals. University of Nebraska. <http://johnmacfarlane.net/cloudy.pdf>.

Might do better

- MacFarlane, John. 2011. Epistemic modals are assessment sensitive. In Andy Egan & Brian Weatherson (eds.), *Epistemic modality*, chap. 5, 144–178. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199591596.003.0006>.
- MacFarlane, John. 2014. *Assessment-sensitivity: Relative truth and its applications*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199682751.001.0001>.
- Moss, Sarah. 2015. On the semantics and pragmatics of epistemic vocabulary. *Semantics and Pragmatics* 8(5). 1–81. <https://doi.org/10.3765/sp.8.5>.
- Ninan, Dilip. 2018. Relational semantics and domain semantics for epistemic modals. *Journal of Philosophical Logic* 47(1). 1–16. <https://doi.org/10.1007/s10992-016-9414-x>.
- Plunkett, David & Timothy Sundell. 2013. Disagreement and the semantics of normative and evaluative terms. *Philosophers' Imprint* 13(23). 1–37. <https://doi.org/2027/spo.3521354.0013.023>.
- Portner, Paul. 2009. *Modality*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780199292424.001.0001>.
- Roberts, Craige. 1996–2012. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6). 1–69. <https://doi.org/10.3765/sp.5.6>.
- Roberts, Craige. 2003. Uniqueness in definite noun phrases. *Linguistics and Philosophy* 26(3). 287–350. <http://www.jstor.org/stable/25001888>.
- Roberts, Craige. 2017. *Agreeing and assessing: Epistemic modals and the question under discussion*. Oslo workshop on ‘non-at-issue meaning and Information Structure’.
- Roberts, Craige. 2023. The indexical character of epistemic modality. *Linguistics and Philosophy* 46(5). 1219–1267. <https://doi.org/10.1007/s10988-023-09384-3>.
- Schaffer, Jonathan. 2011. Perspective in taste predicates and epistemic modals. In Andy Egan & Brian Weatherson (eds.), *Epistemic modality*, chap. 6, 179–226. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199591596.003.0007>.
- Schaffer, Jonathan & Zoltán Gendler Szabó. 2014. Epistemic comparativism: A contextualist semantics for knowledge ascriptions. *Philosophical Studies* 168. 491–543. <https://doi.org/10.1007/s11098-013-0141-7>.
- Simons, Mandy. 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117(6). 1034–1056. <https://doi.org/10.1016/j.lingua.2006.05.006>.

- Spencer, Jack. 2016. Disagreement and attitudinal relativism. *Mind* 125(498). 511–539. <https://doi.org/10.1093/mind/fzv150>.
- Stephenson, Tamina. 2007a. Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy* 30(4). 487–525. <https://doi.org/10.1007/s10988-008-9023-4>.
- Stephenson, Tamina. 2007b. *Towards a theory of subjective meaning*. Cambridge, MA: Massachusetts Institute of Technology dissertation. <https://doi.org/1721.1/41695>.
- Sundell, Timothy. 2011. Disagreements about taste. *Philosophical Studies* 155(2). 267–288. <https://doi.org/10.1007/s11098-010-9572-6>.
- Swanson, Eric. 2006. *Interactions with context*. Cambridge, MA: Massachusetts Institute of Technology dissertation. <https://doi.org/1721.1/37356>.
- Veltman, Frank. 1996. Defaults in update semantics. *Journal of Philosophical Logic* 25(3). 221–261. <https://doi.org/10.1007/BF00248150>.
- Werner, Tom. 2006. Future and non-future modal sentences. *Natural Language Semantics* 14(3). 235–255. <https://doi.org/10.1007/s11050-006-9001-8>.
- Willer, Malte. 2013. Dynamics of epistemic modality. *Philosophical Review* 122(1). 45–92. <https://doi.org/10.1215/00318108-1728714>.
- Yalcin, Seth. 2007. Epistemic modals. *Mind* 116(464). 983–1026. <https://doi.org/10.1093/mind/fzm983>.
- Yalcin, Seth. 2010. Probability operators. *Philosophy Compass* 5(11). 916–937. <https://doi.org/10.1111/j.1747-9991.2010.00360.x>.
- Yalcin, Seth. 2011. Nonfactualism about epistemic modality. In Andy Egan & Brian Weatherson (eds.), *Epistemic modality*, chap. 10, 295–332. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199591596.003.0011>.
- Yalcin, Seth. 2015. Epistemic modality *de re*. *Ergo* 2(19). 475–527. <https://doi.org/10.3998/ergo.12405314.0002.019>.
- Yanovich, Igor. 2014. Standard contextualism strikes back. *Journal of Semantics* 31(1). 67–114. <https://doi.org/10.1093/jos/ffso22>.
- Zondervan, Arjen, Luisa Meroni & Andrea Gualmini. 2008. Experiments on the role of the question under discussion for ambiguity resolution and implicature computation in adults. *Semantics and Linguistic Theory (SALT)* 18. 765–777. <https://doi.org/10.3765/salt.v18io.2486>.

Might do better

Bob Beddor
Department of Philosophy
University of Florida
r.beddor@ufl.edu

Andy Egan
Department of Philosophy
Rutgers University
andyegan@rutgers.edu