

New horizons for a theory of epistemic modals

Justin Khoo & Jonathan Phillips

Abstract

The centerpiece of the recent debate over the semantics and pragmatics of epistemic modals concerns intuitions about cross-contextual truth-value assessments. In this paper, we advocate for a change in the focus of this debate and a new way of evaluating theories of epistemic modals. We argue that intuitions about cross-contextual truth-value assessments are ill-suited for deciding between the major theories of epistemic modals. We then go on to develop a new test which instead exploits the structural differences in the relevant theories by focusing on judgments of when two different epistemic possibility claims should be judged to be compatible. We subject the predictions of existing theories to empirical scrutiny, and find that no existing theory is able to account for the full pattern of observed judgments. As a way of illustrating the theoretical upshot of these results, we conclude by developing a novel theory of epistemic modals that is able to predict the results.

Keywords: epistemic modals, contextualism, relativism, truth.

Word count: 8,164

At the core of the recent debate over the semantics of modals is a tension between modal language and its subject matter. Take, for instance, epistemic modal auxiliaries like “might,” “could,” “must” and “probably.”¹ On the face of it, a bare epistemic modal sentence like (1) seems to state that it is epistemically possible that the keys are in the drawer:

(1) The keys might be in the drawer.

However, whether a proposition is epistemically possible is a relative matter: p may be epistemically possible relative to some bodies of evidence and not others. Thus, we are faced with a mismatch between the subject matter of epistemic modality and the language we use to talk about it.

How we resolve this mismatch is the subject of a recent debate over whether the semantic/pragmatic behavior of epistemic modals is best captured within a contextualist or relativist theory. Modal contextualism resolves the mismatch by proposing that, contrary to appearances, epistemic modal sentences like (1) in fact express propositions that are about particular bodies of evidence. So, for instance, an epistemic possibility sentence like (1) might express that the keys

¹There are other flavors of modals (deontic, metaphysical, bouletic), but we won't discuss them here.

being in the drawer is compatible with (read: not ruled out by) the evidence that John has. Contextualism holds that *what* body of evidence a particular epistemic modal claim is about depends on the context in which it is uttered (Hacking (1967); Kratzer (1981, 1991); DeRose (1991); Stanley (2005); Dowell (2011)). So, uttered in one context, (1) expresses the claim that John’s evidence does not rule out the keys being in the drawer; uttered in a different context, it expresses the claim that B’s evidence does not rule out the keys being in the drawer (and so on).

By contrast, modal relativism resolves the mismatch by proposing that the truth or falsity of an epistemic modal claim varies by the context in which it is assessed (cf. MacFarlane (2011, 2014)). Epistemic modal sentences like (1) are not about particular bodies of evidence; rather, they express propositions whose truth values vary across bodies of evidence.² The truth value of a modal claim depends on the truth of the expressed proposition relative to the body of evidence that is relevant at the context in which that claim is assessed.

The debate between contextualist and relativist theories has tended to focus on cross-contextual assessments of speech acts involving bare epistemic modal sentences like (1). These are cases where a more-informed eavesdropper secretly listens in on a conversation in which a lesser-informed utterer utters a bare epistemic modal sentence like (1). An early claim in the literature was that if someone who doesn’t know where the keys are says (1), and if an eavesdropper knows that the keys are not in the drawer, the eavesdropper will correctly judge of the speaker that she said something false (Egan *et al.* (2005); Egan (2007); Stephenson (2007); MacFarlane (2011)). Since this is a natural prediction of relativism, but not contextualism, this observation was put forward as evidence that some form of relativism about epistemic modals was correct.

Recently, however, many non-relativists have pushed back against this line of argument. These

²This distinction glosses over two others which are relevant. The first is between what we have called modal relativism and what MacFarlane (2009) calls “non-indexical contextualism.” The latter agrees with the former that the truth value of the proposition expressed by an epistemic modal sentence is relative to a body of evidence. But it disagrees about which body of evidence matters to the truth of the claim made when someone utters an epistemic modal sentence. According to non-indexical contextualism, the truth value of a modal claim depends on the truth of the expressed proposition relative to the body of evidence relevant at the context of utterance; whereas for relativism, the truth value of a modal claim depends on the truth of the expressed proposition relative to the body of evidence relevant at the context in which that claim is assessed. The second distinction is between what we are calling modal relativism and what we might call “modal judge-dependence.” The latter is relativist-like in that it holds that the truth value of a proposition expressed by an epistemic modal sentence is relative to a judge (cf. Egan *et al.* (2005); Egan (2007); Stephenson (2007)). This allows for modal sentences to be more fine-grained than factual sentences (in a sense to be clarified below). We will come back to this view in §4.

theorists argue that intuitions about such cases are much more complex (Knobe & Yalcin (2014)), and in fact favor contextualism (Dowell (2011)) or expressivism (Yalcin (2011)).

Our aim in this paper is to shift the perspective on this debate and introduce a new way of evaluating contextualist and relativist theories, not just about epistemic modals, but across a range of domains.³ Starting with Knobe & Yalcin (2014)’s empirical results, which have been taken to pose a challenge to relativism, we argue that, in fact, the results do not distinguish between the two theories at all (§1). While there are ways to improve the strategy of testing ordinary intuitions of cross-contextual assessments, we instead propose a different way to distinguish the theories, one which focuses entirely on structural differences between them (§2). Then, in §3, we report the results of a new empirical study, which reveals data that neither existing contextualist nor relativist theories predict. This motivates the search for a new kind of theory, which we sketch in §4. Our theory combines two ideas: the first that the propositions expressed by epistemic modal sentences sensitive to indices that are finer-grained than possible worlds, and the second is a contextualist theory of truth. We show how our theory can predict our experimental results from §3. We conclude with a brief discussion of some broader lessons of our new approach to thinking about the debate between contextualism and relativism.

1 Eavesdroppers

Let’s begin with a simple semantic theory for epistemic modals, which is neutral between contextualist and relativist interpretations:

SIMPLE SEMANTICS

$$\llbracket \text{‘might } p \text{’} \rrbracket^{f,w} = 1 \text{ iff } \exists w' \in f(w): \llbracket p \rrbracket^{f,w'} = 1.$$

A brief note on our choice of terminology. ‘ $\llbracket \ \rrbracket$ ’ is our interpretation function—it maps expressions of English to their extensions relative to points of evaluation. We are assuming for now that a point

³The debate over semantic/pragmatic frameworks extends to a wide range of philosophically interesting natural language expressions—normative expressions (both epistemic and moral), predicates of personal taste, gradable adjectives, conditionals, to name a few. Although our discussion focuses entirely on epistemic modals, we think the methods we develop here extend quite naturally to these other expressions, and thus may be useful in assessing a wide range of interesting issues.

of evaluation is pair of a domain-fixing function f , and a world w . The domain-fixing function maps worlds to sets of worlds, which is the domain of the modal—it is our way of modeling a body of evidence (or, more precisely, its informational content). Finally, notice that we are assuming that modal sentences like (1) are syntactically composed of a modal, ‘might,’ scoping over a sentence ‘p’—we make this assumption for convenience and because it is widely assumed in the literature.

The difference between contextualism and relativism can now be stated in terms of which post-semantic theory we add to SIMPLE SEMANTICS. According to the contextualist, we need to add one additional parameter to our point of evaluation—a context of utterance. The context of utterance is what (post-semantically) fixes the value of the domain-fixing function and world relative to which the truth of an epistemic modal sentence, as uttered in that context, is determined:

CONTEXTUALIST POST-SEMANTICS

‘might p’ is true as uttered in c iff $\llbracket \text{‘might p’} \rrbracket^{c, f_c, w_c} = 1$.

Now, we can define the *content* of an utterance of ‘might p’ in c as the possible-worlds proposition $\{w : \llbracket \text{‘might p’} \rrbracket^{c, f_c, w} = 1\}$. This proposition is *about* the body of evidence f_c because its truth (at w) depends (entirely) on how things are with $f_c(w)$.

By contrast, the relativist offers an alternative post-semantics. To begin, the relativist proposes adding *two* parameters to our points of evaluation—a context of utterance c_U and a context of assessment c_A . This allows the relativist to post-semantically distinguish the roles of the context of utterance and assessment. For an epistemic modal sentence uttered in c_U and assessed in c_A , the context of utterance fixes the world and the context of assessment fixes the value of the domain-fixing function:

RELATIVIST POST-SEMANTICS

‘might p’ is true as uttered in c_U and assessed in c_A iff $\llbracket \text{‘might p’} \rrbracket^{c_U, f_{c_A}, w_{c_U}} = 1$.

According to the relativist, the content of an utterance of ‘might p’ in c_U cannot be modeled as a possible-worlds proposition. This is because the truth of ‘might p’ as uttered in c_U may vary depending on the context in which it is assessed. Instead, its content could be modeled as a set of pairs of a world and domain-fixing function as follows: $\{\langle w, f \rangle : \llbracket \text{‘might p’} \rrbracket^{c_U, f, w} = 1\}$. This

proposition is not about any particular body of evidence—rather its truth value varies across bodies of evidence.

With this more precise characterization of a contextualist and relativist implementation of the SIMPLE SEMANTICS in hand, let's turn now to how eavesdropper cases have been used to motivate relativism over contextualism.

In an eavesdropper case, a better-informed eavesdropper assesses as false an epistemic possibility sentence uttered by someone who does not know whether its prejacent is false (cf. Egan *et al.* (2005); Stephenson (2007); MacFarlane (2011)). Here is one such case (from Knobe & Yalcin (2014)):

Fat Tony. 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 word gets out about his apparent death. The next evening, from his safehouse, Fat Tony watches a panel of experts on the news discussing the question of whether he is dead.

Expert A has had a good look at the evidence found at the scene. “Fat Tony is dead”, he says. Expert B has also had a good look at the evidence, but his assessment is more cautious. B says,

(2) Fat Tony might be dead.

In this case, Expert B does not know whether Fat Tony is dead, even though the case clearly states that Fat Tony is not dead. So, is what Expert B said true in this case? According to relativism, whether what Expert B said is true may depend on the information of the person assessing the truth of Expert B's claim. In particular, many relativists have claimed that, in a case like this, people reading the vignette **Fat Tony** would agree that what Expert B said is false. Suppose for a moment that this is in fact the observation to be explained. Relativism is motivated by this observation because it, but not contextualism, seems able to predict it.

Here is why contextualism has trouble predicting that what Expert B said is false in **Fat Tony**. Given CONTEXTUALIST POST-SEMANTICS, whether ‘might p’ is true as uttered in *c* depends entirely on whether *p* is compatible with the evidence determined by *c*. This view then has trouble predicting that an eavesdropper who is not a part of *c* would correctly assess the proposition expressed by ‘might p’ as uttered in *c* as false merely because *they* (the eavesdropper) know that *p* is false. The point is not that it is impossible for a contextualist theory to predict this result, but that doing so would require building the eavesdropper's evidence into the evidence determined by

c , and this may be implausible in many cases. One looming challenge (pressed aptly by MacFarlane (2011)) is that building the eavesdropper's evidence into the evidence determined by c risks making the truth value of many epistemic possibility sentences implausibly unknowable, and thus predicts that such sentences should be less easily assertable than they otherwise seem.

By contrast, RELATIVIST POST-SEMANTICS easily predicts the alleged eavesdropper judgment. On that theory, \lceil might p \rceil is true as uttered in c_U and assessed in c_A if and only if p is compatible with the evidence determined by c_A . On this kind of relativism, the truth of utterances of (or propositions expressed by) epistemic modal sentences are relative to contexts of assessment, and so the assessor's evidence may determine the truth value of the epistemic modal claim. Relativism thus distinguishes the assertability of \lceil might p \rceil (which goes by its truth value as uttered in the context of utterance, relative to the context of utterance) from its truth value as assessed at various other contexts. It is this feature that allows relativism to avoid the dilemma facing the contextualist above regarding eavesdropping cases.

More recently, however, this argument in favor of relativism over contextualism has been contested on the grounds that the intuitions the relativist appeals to are simply not widely shared (see Dowell (2011); Knobe & Yalcin (2014)). In what follows, we discuss the results of Knobe and Yalcin, who empirically tested the relativists' claim about eavesdropper scenarios. In their first study, they asked ordinary native speakers of English to read the vignette **Fat Tony** and then asked participants whether they agreed or disagreed with the following claims:

- (nonmodal-true) What Expert A said is true.
- (nonmodal-false) What Expert A said is false.
- (modal-true) What Expert B said is true.
- (modal-false) What Expert B said is false.

The results from this experiment are summarized in the following graph:

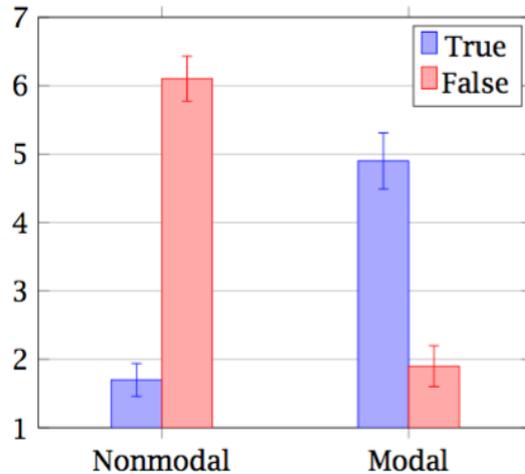


Figure 1: Knobe & Yalcin data.

As is easily seen in the graph, participants tended to think what Expert A said was false, and what Expert B said was true. This is exactly the opposite of what relativists have typically claimed about eavesdropper cases—relativists have generally thought in such cases that assessors will use their own evidence (which includes the information that Fat Tony is alive) in assessing what Expert B said, and thus conclude that what Expert B said is false (Egan *et al.* (2005); Egan (2007, 2010); MacFarlane (2011, 2014); “cloudy contextualists” like von Fintel & Gillies (2011) also aim to predict this result).

Importantly, Knobe and Yalcin do not claim that these data are evidence that relativism is false. Rather, they only claim that they provide reason to think that one kind of evidence that has typically been marshaled in favor of relativism (i.e., eavesdropper scenarios) doesn’t actually seem to support it. This is an important point, and one that bears reemphasizing here.

To see how the results of this experiment are compatible with relativism, we have to appreciate the upshot of MacFarlane (2011)’s discussion of the “challenge of lost information” for relativism (raised by Dietz (2008)). Suppose yesterday I proved Theorem φ but that today I realize that I have lost all memory of the proof (and hence now do not know whether Theorem X is true). Suppose today someone tells me that before I lost my memory I was heard saying,

- (3) Theorem φ must be true.

It seems that today I should judge that I am not sure whether my earlier claim was true. Crucially, though, it may seem that if relativism is true then my saying today (about my past utterance of (3)),

(4) That's false.

should be correct. After all, the information now available to me when I'm assessing (3) does not entail that Theorem φ is true. However, as Macfarlane notes, "This seems bizarre. Intuitively, I don't have warrant to pronounce on the falsity of claims made by my better-informed past self, even when these claims contain epistemic modals." (MacFarlane (2011): 174). Macfarlane's response is to combine RELATIVIST POST-SEMANTICS with an account of what f_{c_A} may be, for some c_A s. Initially, one might have thought that the evidence determined by c_A must just be the evidence available to the assessor(s) in that context. If this were the case, though, the theory would predict the wrong results in the "information loss" cases and fail to predict Knobe and Yalcin's data in the **Fat Tony** case. But Macfarlane correctly points out that this is a defeasible commitment of relativism. It is compatible with RELATIVIST POST-SEMANTICS that the evidence determined by c_A is the evidence that is relevant in c_U . Making this move allows the relativist to hold that the relevant information in the context of assessment *may* (but need not always) be the information available to the agent at c_A ; in some cases, the relevant evidence at the context of assessment will be the evidence available to someone else (perhaps the speaker at the context of utterance).⁴

Along these lines, it is open for Macfarlane to account for the results of Knobe and Yalcin's first experiment by holding that participants tended to judge that the relevant information (in their context) was the evidence available to Expert B (the evidence at the docks), which suggested that Fat Tony had died. Which information participants took to be relevant when they were assessing these claims was simply not tested, and thus these studies do not provide direct evidence against relativism.

This point tempers the importance of Knobe and Yalcin's data for the contextualism/relativism debate. But we actually think it points toward a more radical concern with this approach. In

⁴This theory is still different from contextualism because it allows that the truth of \lceil might p \rceil as uttered at c_U will still vary across contexts of assessment, owing to different evidence being relevant at those contexts.

particular, we think that even if the results of the studies had confirmed the intuitions reported by many relativists, this would not have been evidence *against* contextualism. This is because, in that case, the contextualist could reply that her theory is one on which assessments about the truth/falsity of an epistemic modal claim can be used as evidence about which claim was made by the original utterance (as in Dowell (2013)’s response about better-informed assessments of information-sensitive deontic modal claims). In other words, the contextualist could hold that, even if the relativist’s intuitions were empirically confirmed, these results simply reveal that ordinary people sometimes use their own evidence to make up their minds about what claim was made in the original utterance context.

Since the evidence from eavesdropping scenarios like the one used in Knobe and Yalcin’s study is compatible with both contextualism and relativism, regardless of how the results might have turned out, we conclude that the actual results of their study is not helpful in deciding between the two theories.⁵

Let’s recap the existing dialectic. Relativistic theories differ from contextualist theories in their post-semantics. One important prediction that distinguishes the theories is that relativism allows the truth of an epistemic modal claim made in some context to vary across assessment contexts in which different evidence is relevant. Thus, the intuition that an eavesdropper who knows $\neg p$ would be correct to judge that a lesser-informed speaker who utters ‘‘might p’’ to have said something false seems to provide a challenge to contextualism, and a point in favor of relativism. Knobe and Yalcin provided empirical evidence that this intuition is not shared by ordinary individuals, for at least some cases. We argued that this outcome is compatible with a flexible version of relativism, one that Macfarlane is independently motivated to adopt anyway. But, more surprisingly, we also argued that the ‘‘relativist friendly’’ eavesdropper intuitions are also compatible with a version of contextualism that allows that facts about the context of assessment may be evidentially relevant to what claim the original speaker made. Therefore, we think that exploring the truth value intuitions of eavesdroppers (or data about agreement/disagreement with eavesdropper assessments) is not a

⁵This is not to deny that their studies were important. After all, as noted above, a widely reported intuition in the literature was that an eavesdropper who knows $\neg p$ would judge that a lesser-informed speaker who utters ‘‘might p’’ to have said something false. And *this* intuition, at least in the kinds of case Knobe and Yalcin discuss, seems not to be shared by ordinary speakers.

promising way to empirically distinguish between contextualism and relativism.

2 Structural differences

In this section, we propose a new way to test the predictions of contextualism and relativism. Rather than focus on the truth value of a single claim, we will focus on the question of whether two claims are incompatible by asking people how much they agree or disagree with the statement:

(Q) At least one of X's or Y's claims must be false.

Agreeing with (Q) is to judge that X and Y's claims are incompatible, while disagreeing with (Q) is to judge that X and Y's claims are not incompatible (whether this *just is* to judge that the claims are compatible is an issue we set aside).

There are two targets of incompatibility judgments that we will explore. The first is the incompatibility of two modal claims. The second is the incompatibility of two assessments of a (single) modal claim. In the next two sub-sections, we discuss each type of case, and what contextualism and relativism predict about it.

2.1 Modal Utterances

In the **Modal Utterances** case, two speakers (who we will generically name 'X' and 'Y') are in different contexts (i.e., they are not talking to each other). X has evidence that is compatible with the proposition p , whereas Y has evidence that entails $\neg p$. X utters \lceil might p \rceil in his context, while Y utters \lceil not-might p \rceil . Contextualist and relativist theories predict different results for (Q) in this kind of case. A contextualist theory can allow for the possibility that X and Y's claims are compatible (thus predicting (Q) is false), and can also allow for the possibility that X and Y's claims are incompatible (thus predicting that (Q) is true). By contrast, any relativist theory will predict that (Q) is true. This is because relativism predicts that (relative to the context in which the question is asked and answered), at least one of X and Y's claims must be false.

Here is why contextualist and relativist theories differ in this way with regards to (Q). According to contextualism, X asserts the proposition that p is compatible with the evidence relevant in her

context f_x , while Y asserts the proposition that p is not compatible with the evidence relevant in *her* context f_y . Now, it could be that p is compatible with the evidence in X’s context and not compatible with the evidence in Y’s context – in that case, X and Y would have made compatible claims. But it could also be that X and Y’s evidence is sufficiently similar, so that it can’t be that p is compatible with X’s evidence and not compatible with Y’s – in that case, X and Y would have made incompatible claims. Thus, contextualism can predict either that (Q) is true or that (Q) is false, depending on how we understand the evidence in X and Y’s contexts.

By contrast, relativism entails that X and Y’s claims are incompatible. According to relativism, what Y asserts (call it $\diamond p$) just is the negation of what X asserts ($\neg\diamond p$).⁶ But then, relative to any context of assessment, it can’t be that both of their claims are true. To see why, recall that RELATIVIST POST-SEMANTICS and SIMPLE SEMANTICS together entail the following:

$\diamond p$ is true at a world w as assessed at a context of assessment c_A iff $f_{c_A}(w)$ is compatible with p .

$\neg\diamond p$ is true at a world w as assessed at a context of assessment c_A iff $f_{c_A}(w)$ is not compatible with p .

But of course, for no world and context of assessment c_A is $f_{c_A}(w)$ both compatible and not compatible with p . Therefore, relative to any context of assessment, relativism predicts that (Q) is true.

Since both (modal) contextualism and relativism are theories about the semantics and pragmatics of epistemic natural language modals, we expect these theories to account for the linguistic intuitions of native speakers of the language under consideration (in our case, English). Therefore, the differences in the theories’ predictions about (Q) provide a way to empirically test these two theories. One way to do this is to compare the pattern of intuitions about (Q) in a **Modal Utterances** case (like the one above) with a **Non-Modal Utterances** case, which is exactly like the former except that the two characters make conflicting factual claims (i.e., X utters p and Y utters not- p). In the non-modal version, we expect that, native speakers will tend to agree with

⁶As above, we might model these propositions as sets of world, domain-fixing-function pairs: $\diamond p = \{\langle w, f \rangle : \llbracket \text{might } p \rrbracket^{c_U, c_A, f, w} = 1\}$; $\neg\diamond p = \{\langle w, f \rangle : \llbracket \text{might } p \rrbracket^{c_U, c_A, f, w} = 0\}$.

(Q). Since relativism predicts a similar pattern for (Q) in the modal version, relativism should thus predict that native speakers will agree with (Q) to the same extent in **Modal Utterances** as in **Non-Modal Utterances**. By contrast, contextualism allows for the possibility of a difference in level of agreement with (Q) between **Modal Utterances** and **Non-Modal Utterances**. Thus, finding a significant difference in the pattern of answers to (Q) between **Modal Utterances** and **Non-Modal Utterances** would be evidence for contextualism and against relativism; finding no difference between the two cases would be evidence compatible with both theories.⁷

2.2 Modal Assessments

In the **Modal Assessments** case, a third party (who we generically name ‘Z’), utters \lceil might p \rceil . Two individuals, X and Y again, who are in completely different contexts from Z (and each other), overhear Z’s utterance. As before, X’s evidence is compatible with the proposition p , while Y’s evidence entails that $\neg p$. X utters (5), while Y utters (6):

(5) What Z said is true.

(6) What Z said is false.

In this case, both contextualism and relativism predict that (Q) is true: that is, both theories predict that at least one of X or Y’s claims in this scenario must be false.

To see why, we first need a semantics for speech reports (“what Z said”) and for the truth predicate. Regarding the former, we make the orthodox assumption that “What Z said” refers to the proposition Z asserted, as follows (we will discuss this assumption in more detail below):

$$\llbracket \text{“What Z said”} \rrbracket^{c,f,w} = \text{the proposition Z asserted at } w.$$

Regarding the latter, we will adopt a deflationist semantics for “is true”/“is false” on which “What Z said is true” is necessarily equivalent to the proposition referred to by “What Z said” and “What Z said is false” is necessarily equivalent to the negation of the proposition referred to by “What Z

⁷After writing this paper, we learned of Katz & Salerno (2016), who ran some studies similar to our **Modal Utterances** case. Katz and Salerno did not, however, test incompatibility judgements in a **Modal Assessment** kind of case. As will become evident, judgments in the modal assessments cases prove to be much more problematic for existing theories and motivate the larger revision which we pursue here.

said.” This yields:

DEFLATIONARY CONTEXTUALISM:

$$\llbracket \text{“What Z said is true”} \rrbracket^c = \{w: \llbracket \text{“What Z said”} \rrbracket^{c,f_c,w_c} \text{ is true at } w\}$$

DEFLATIONARY RELATIVISM:

$$\llbracket \text{“What Z said is true”} \rrbracket^c = \{\langle w, f \rangle: \llbracket \text{“What Z said”} \rrbracket^{c,f,w_c} \text{ is true at } \langle w, f \rangle\}$$

With these assumptions in hand, we can show why both contextualism and relativism answer “yes” to (Q) in the **Modal Assessments** case.

According to contextualism, when Z says, \lceil might $p \rceil$, he asserts that p is compatible with the evidence in his (Z’s) context. By DEFLATIONARY CONTEXTUALISM, when X says, “What Z said is true,” he asserts the same proposition Z did, and when Y says, “What Z said is false,” he asserts the negation of this proposition. Since no proposition can be both true and false, it follows that at least one of X or Y’s claims must be false.

According to relativism, when Z says, \lceil might $p \rceil$, he asserts $\diamond p$ (recall, the truth value of this proposition is relative to a domain function). By DEFLATIONARY RELATIVISM, when X says, “What Z said is true,” he asserts the same proposition Z did, and when Y says, “What Z said is false,” he asserts the negation of this proposition. Since relative to any context of assessment, no proposition can be both true and false, it follows that at least one of X or Y’s claims must be false.⁸

As above, we assume that in a **Non-Modal Assessments** case (one in which Z utters the non-modal sentence p but is otherwise just like the modal case described above), native speakers will tend to agree with (Q). Therefore, given that both contextualism and relativism predict (Q) is true, given our other assumptions (that “what Z said” refers to the proposition Z asserted and the deflationary semantics for the truth/falsity predicate), both predict no difference in speakers’ pattern of agreement with (Q) in **Modal Assessments** and **Non-Modal Assessments** minimal pairs. Thus, finding a significant difference in the pattern of answers to (Q) between modal and non-modal **Assessments** cases would be evidence against *both* contextualism and relativism (given

⁸This is so even though it may be that X’s claim is true as assessed at X’s context and Y’s claim as assessed at Y’s context. One of the key features of relativism is that it predicts that our judgments should mimic objectivism, and it does this by the fact that, relative to any single context of assessment, at least one of their claims must be false.

these assumptions, which we will discuss in more detail below); finding no difference would be evidence compatible with both theories.

We summarize these predictions of contextualist and relativist theories as follows:

	Contextualism	Relativism
Modal Utterances	Q: ✓ / ✗	Q: ✓
Modal Assessments	Q: ✓	Q: ✓

It is an empirical question which, if any, of these predictions of contextualism and relativism are accurate. That is, we do not yet know whether ordinary speakers of English will judge the claims made by X and Y in these situations to be incompatible. In the next section, we explore this question.⁹

3 Testing the structural differences

3.1 Participants

To test the pattern of predictions discussed above, we conducted an experiment. Two hundred and forty participants ($M_{age} = 37.93$, $SD_{age} = 12.11$, 49% female) were recruited through Amazon Mechanical Turk (www.mturk.com).

3.2 Design

All of the participants read the following background information:

Background: The police are on the trail of Fat Tony, a local mobster. This morning, they learn of a rumor that Fat Tony has died at the docks.

The Chief of the Police assigns Inspector A to examine the evidence at the docks. Meanwhile, the District Attorney assigns Inspector B to review the footage from the security camera at the docks.

⁹We want to emphasize that we are comparing the overall *pattern* of answers between the two cases. We are not suggesting that any particular native speaker answering “no” to (Q) is evidence *by itself* against relativism (or for contextualism). Obviously, relativism is compatible with the possibility of inattention, time-constraints, and misremembering relevant details, all of which may be the actual cause of any particular person’s behavior (their registering some judgment about some question). That is why it is patterns of responses that matter to assessing contextualism and relativism.

Participants were randomly assigned to either a **Utterances** or a **Assessments** case. In each of these cases, they were also randomly assigned to one of three different conditions, which varied whether the statement being evaluated involved an epistemic **Modal** claim, a **Non-Modal** claim, or an **Indexical** claim. To illustrate, in the **Modal Utterances** case, the background continued as follows:¹⁰

Modal Utterances: Inspector A takes a good look at the evidence down by the docks, and concludes that it suggests, but does not prove, that Fat Tony died at the docks. The Chief calls Inspector A at the docks and asks him, “What have you found?”

Inspector A replies, “Fat Tony could have died at the docks.”

Meanwhile, Inspector B reviews the security camera footage and concludes that the footage proves that Fat Tony did **not** die at the docks. The District Attorney calls Inspector B and asks him, “What have you found?”

Inspector B replies, “Fat Tony couldn’t have died at the docks.”

By contrast, in the **Modal Assessments** vignette, the two inspectors made two different claims about the truth of the Chief’s utterance:

Modal Assessments: Inspector A takes a good look at the evidence down by the docks, and concludes that it suggests, but does not prove, that Fat Tony died at the docks. Afterwards, he goes home. That evening, Inspector A and his wife watch the Chief of Police talking with reporters on TV. The reporters on the news ask the Chief what his investigation had found.

The Chief tells the reporters: “Fat Tony could have died at the docks.”

Inspector A’s wife knows that Inspector A was examining the evidence at the docks and so she asks him, “Is that right?”

Inspector A replies, “What the Chief said is true.”

Meanwhile, Inspector B reviews the security camera footage and concludes that the footage proves that Fat Tony did not die at the docks. That evening he watches the same TV broadcast with his wife, and they also hear the Chief tell the reporters, “Fat Tony could have died at the docks.”

Inspector B’s wife knows that Inspector B was examining the evidence at the docks and so she asks him, “Is that right?”

Inspector B replies, “What the Chief said is false.”

¹⁰We changed the target sentence from a modal claim about the present (“Fat Tony could be dead”) to one about the past (“Fat Tony could have died at the docks”) to remove the possibility of thinking that Tony didn’t die at the docks but instead died sometime thereafter. However, doing so opens up the possibility that some participants interpret “could have” metaphysically, rather than epistemically (as we intend). However, we do not think this interpretive possibility interferes with our results, since the metaphysical interpretation is clearly deviant in the context in which the sentence is uttered (the Inspector is asked to report what he has found in the course of investigating whether Fat Tony died at the docks).

After reading, participants were reminded that the inspectors had made two different claims and were asked whether they agreed or disagreed that “At least one of the inspector’s claims must be false.” Participants rated their agreement on a scale from 1 (“Completely Disagree”) to 7 (“Completely Agree”).

For comparison, other participants instead evaluated a Non-modal claim. These cases were identical to the preceding ones except that the Inspectors’/Chief’s claim(s) did not include the epistemic modal, and thus read: “Fat Tony died [did not die] at the docks.”

After answering this question, all participants in both the Modal and Non-modal conditions were asked to make a judgment about what was more relevant in Inspector A’s conversation and, then separately, in Inspector B’s conversation, which allowed us to ask whether they tracked the differences across these two conversational contexts. In both cases, participants responded by selecting which of the following two options was more relevant in each conversation:

- What the evidence at the docks reveals about Fat Tony.
- What the security camera footage reveals about Fat Tony.

In addition to the Modal and Non-modal conditions, we also included a condition in which the relevant statement involved an indexical claim, which provided a separate test of how participants understood the meaning of “what the Chief said.” In the **Utterances** case, **Background** continued as follows:

Indexical Utterances: Inspector A takes a good look at the evidence down by the docks, and concludes that it suggests, but does not prove, that Fat Tony died at the docks. Later that evening, Inspector A gets a call from the Chief. The Chief knows that certificates of appreciation are being given to officers who have served on the police force for at least twenty years, so he asks Inspector A, “How long have you served on the police force?”

Inspector A replies, “I have served on the police force for twenty years.”

Meanwhile, Inspector B reviews the security camera footage and concludes that the footage proves that Fat Tony did **not** die at the docks. Later that evening, Inspector B gets a call from the District Attorney. The District Attorney also knows that certificates of appreciation are being given to officers who have served on the police force for at least twenty years, so he asks Inspector B, “How long have you served on the police force?”

Inspector B replies, “I have not served on the police force for twenty years.”

By contrast, in the indexical version of **Assessments**, the two inspectors made two different claims about the truth of the Chief's utterance:

Indexical Assessments: Inspector A takes a good look at the evidence down by the docks, and concludes that it suggests, but does not prove, that Fat Tony died at the docks. Afterwards, he goes home. That evening, Inspector A and his wife watch the Chief of Police talking with reporters on TV. The reporter on the news knows that certificates of appreciation are being given to officers who have served on the police force for at least twenty years, so she asks the Chief, "How long have you served on the police force?"

The Chief tells the reporters: "I have served on the police force for twenty years."

Inspector A's wife knows that Inspector A is on the police force, and so she asks him, "Is that right?"

Inspector A replies, "What the Chief said is true."

Meanwhile, Inspector B reviews the security camera footage and concludes that the footage proves that Fat Tony did not die at the docks. That evening he watches the same TV broadcast with his wife, and they also hear the Chief say to the reporter, "I have served on the police force for twenty years."

Inspector B's wife knows that Inspector B was also on the police force, and so she asks him, "Is that right?"

Inspector B replies, "What the Chief said is false."

Participants in these cases were also reminded that the inspectors made two different claims and were asked whether they agreed or disagreed that "At least one of the inspector's claims must be false." Participants again rated their agreement on a scale from 1 ("Completely Disagree") to 7 ("Completely Agree"). We hypothesized that participants would understand "What the Chief said" to pick out the proposition the Chief asserted, *not* something more like the character of the sentence he uttered, and thus that there would be significantly more agreement with "At least one of the inspector's claims must be false" in the **Indexical Assessments** case than in the **Indexical Utterances** case.

3.3 Results

We first analyzed participants' judgments of whether one of the Inspectors' claims must be false in the **Indexical** condition. In the **Indexical Utterances** condition, where Inspector A says, "I have served on the force for more than twenty years," and Inspector B says "I have not served on the force for more than twenty years," participants strongly disagreed that at least one of the Inspectors claims must be false ($M = 2.64$, $SD = 1.96$). However, in the **Indexical Assessments**

condition, where the two Inspectors made conflicting assessments about the Chief’s utterance of “I have served on the police force for twenty years,” participants instead strongly agreed that at least one of the Inspectors claims must be false ($M = 5.67$, $SD = 1.40$), $t(68.69) = -7.82$, $p < .001$, $d = 1.77$, (see Fig. 2).

In brief, this clear pattern suggests that when participants are evaluating the Inspectors’ assessments of “what the Chief said,” they understood the Inspectors to be referring to the content (the proposition asserted) rather than the character of the Chief’s utterance. Had they instead thought the Inspectors were referring to the character of the Chief’s utterance, they would have understood Inspector A’s claim to be equivalent to the claim he would have expressed had he instead asserted, “I have served on the police force for more than twenty years,” and they would have understood Inspector B’s claim to be equivalent to the claim he would have expressed had he instead asserted, “I have not served on the police force for more than twenty years.” But these two claims *are* compatible! Thus, since participants thought that the Inspectors’ claims were incompatible in the **Indexical Assessments** condition, the natural conclusion to draw is that participants understood the Inspectors to be referring to the content of the Chief’s utterance (that is, the proposition he asserted). This is confirmation of our hypothesis that participants understand “What the Chief said” as a locution used to refer to the proposition the Chief asserted.

With this in mind, turn next to participants’ responses in the **Modal** and **Non-modal** conditions. To ensure that participants correctly understood the relevant differences in Inspector A’s and Inspector B’s contexts, we first assessed participants’ judgments of which evidence was most relevant in the two contexts. For both the modal and non-modal conditions, these judgments of relevance confirmed that participants largely tracked the changes in the different contexts: participants found the evidence at the docks to be more relevant in Inspector A’s context, and found the evidence from the security camera to be more relevant in Inspector B’s context, $\chi^2(1) = 74.39$, $p < .001$, $V = .508$.

Finally, we analyzed participants’ compatibility judgments in the Modal and Non-modal conditions with a 2 (Claim: Modal vs. Non-modal) x 2 (Case: Utterances vs. Assessments) ANOVA. We found that participants’ ratings were significantly affected by whether or not the claims in-

volved an epistemic modal, $F(1, 140) = 29.204$, $p < .001$, $\eta_p^2 = .171$, such that they strongly agreed that one of the inspector’s claims must be false when they uttered/assessed a non-modal claim ($M = 5.96$, $SD = 1.33$), but not when they uttered/assessed a modal claim ($M = 4.38$, $SD = 2.10$), $t(119.89) = 5.11$, $p < .001$, $d = .902$. In contrast, we did not observe a significant effect of whether the Inspectors made conflicting utterances or conflicting assessments, $F < 0.5$, and most importantly did not find an interaction effect between these two variables, $F < 1.5$, meaning that the *difference* between the modal and non-modal claims did not significantly differ between the **Assessments** and **Utterances** conditions.

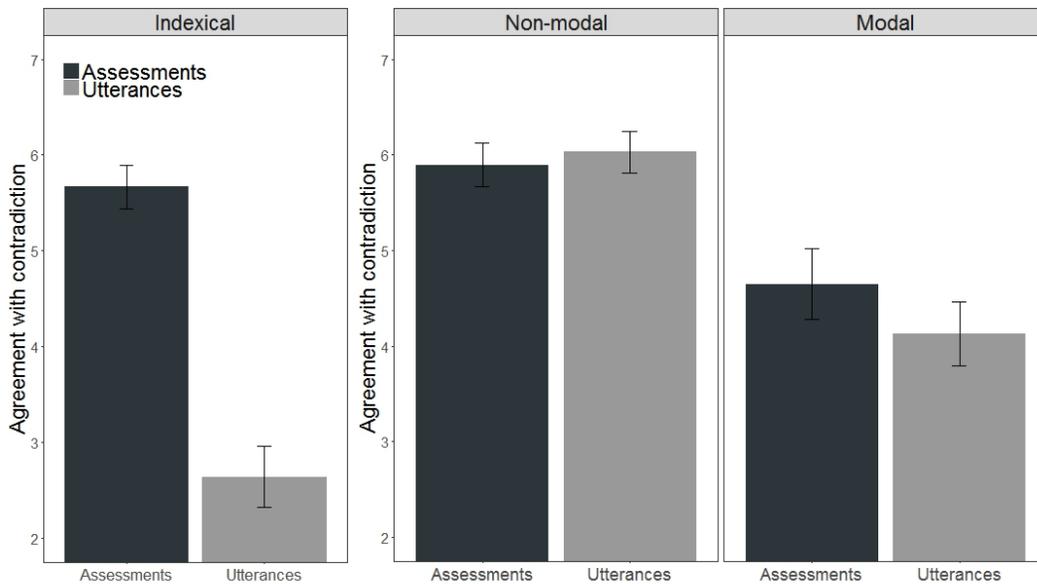


Figure 2: Participants’ mean level of agreement that at least one of the inspectors’ claims must be false. Errors bars indicate $+/- 1SE$.

In brief, what we find is this: In line with contextualism, but not relativism, participants were more inclined to judge that the two modal (vs non-modal) claims were compatible in the **Utterances** condition. At the same time, in conflict with *both* relativism and contextualism, participants were also more inclined to judge that the two assessments were compatible when they concerned a modal (vs non-modal) claim in the **Assessments** condition.

3.4 Discussion

Given the clear predictions made by contextualism and relativism, we take our results (especially the results from the **Assessments** condition) to raise a challenge to both theories of epistemic modals. However, before we move on, we want to pause to consider some alternative hypotheses which might block us from drawing this conclusion.

1. *Is there a better interpretation of participants' incompatibility judgments of epistemic modal claims?*

One natural worry about drawing semantic conclusions from our experiment is the possibility that the ordinary notion of truth and falsity is not the same as the notion that semanticists and philosophers of language care about when they propose theories that predict the truth conditions for natural language sentences. We can think of two possible versions of this worry. The first holds that when participants think about (Q), they are really assessing whether the two inspectors are disagreeing. The second holds that when participants think about (Q), they are really assessing whether at least one of the inspectors was epistemically unjustified in asserting his claim.

(Q) At least of the inspectors' claims must be false.

Take the disagreement worry first. The thought is that participants are confusing a semantic relation of the inspectors' claims (that they can't both be true—call this truth exclusion) with a pragmatic relation: that the inspectors are disagreeing. If this hypothesis were correct, then we should not infer anything about the semantics of epistemic modals from midpoint level agreement with (Q) in **Modal Utterances** and **Modal Assessments**. Instead, participants' midpoint agreement would reflect participants' mixed judgments about whether the two inspectors disagreed, which would be a pragmatic, not semantic, result.

While initially plausible, it turns out that there is a growing body of experimental evidence suggesting that ordinary speakers distinguish between disagreement and falsity. Khoo (2015) found that ordinary speakers are more willing to reject an utterance of \lceil might p \rceil (by saying “No, not p ”) than they are to assess that claim as false. And, building on work by Sarkissian *et al.* (2011),

Khoo & Knobe (2016) found that, in certain cases of moral conflict (where one party says, “That is morally wrong” and another “No, that is not morally wrong”), ordinary speakers agree more with the claim that it was appropriate for one to reject the other’s claim (and agree more with the claim that the two parties disagree) than they do with the thought that at least one of the two claims must be false. Collectively, these results demonstrate that participants typically distinguish between disagreement and falsity. Accordingly, it doesn’t seem promising to propose that they were simply confusing the two in our study.

Next, turn to the justification worry. Remember, this is the suggestion that participants interpret (Q) as really being about whether at least one of the inspectors’ claims must be unjustified. If this hypothesis were correct, then we should not infer anything about the semantics of epistemic modals from midpoint level agreement with (Q) in **Modal Utterances** and **Modal Assessments**. Instead, participants’ midpoint agreement would reveal mixed judgments about whether at least one of the two inspectors’ claims was unjustified, which would be an epistemic, not semantic, result.

Again, we turn to experimental evidence suggesting that this hypothesis is incorrect. Focusing on cases of moral conflict, Sarkissian *et al.* (2011) asked participants (i) about whether at least one of the two speakers’ claims must be incorrect, and (ii) about whether at least one of the two speakers must not have good reason to believe their claim. They found that, participants disagreed with the incorrectness claim more than the unjustified claim. This suggests that ordinary speakers can and do distinguish between truth/falsity and justification/non-justification, and thus our results are also unlikely to have resulted from a simple confusion of the two.

2. *What if we drop the assumption of deflationary truth?*

Although we will argue below that a theory which *can* predict our data is one which gives up on deflationism about the truth/falsity predicates, we think that doing so is not sufficient to save the contextualist and relativist theories currently under consideration. Take contextualism first. According to contextualism, the proposition expressed by ‘might p’ in a context c is $\{w : \exists w' \in f_c(w) : \llbracket p \rrbracket^{c,w} = 1\}$. But now notice that, for any world w , this proposition is either true at w or false at w . So, it cannot be the case that two conflicting truth-assessments are both correct. Dropping

the commitment to deflationism would allow for the possibility of truth/falsity-ascriptions targeting a particular value for some parameter that the truth of $\llbracket \ulcorner \text{might } p \urcorner \rrbracket^c$ is sensitive to. But, since there is no evaluative parameter other than the world for contextualism, this move will not help (since, after all, both inspectors occupy the same world in our scenario). As we will see in §4 below, there will be a way of relaxing exactly this assumption, thus generating an alternative version of contextualism which will be in a position to predict our data.

Turn next to relativism. According to relativism, the proposition expressed by $\ulcorner \text{might } p \urcorner$ in a context c is $\{\langle w, f \rangle : \exists w' \in f(w): \llbracket p \rrbracket^{c,f,w} = 1\}$. This kind of theory might in principle benefit from dropping deflationism about the truth/falsity predicates. The strategy would be to allow that “true” targets a particular modal domain-fixing function f , potentially a different one in different contexts. Hence, on this strategy there would be an interpretation of Inspector A’s utterance of “What the Chief said is true” in which he is saying (roughly) “What the Chief said is true relative to my information,” meaning that “It is true relative to my information that Fat Tony could have died at the docks” (and likewise, with “false,” for Inspector B). If this were right, the two inspectors’ claims could both be true! However, the theory still is unable to predict the possibility of compatible conflicting modal claims (the **Modal Utterances** case). The reason is that the theory continues to predict that, for any two contexts c_1 and c_2 : (invariably) the proposition expressed by $\ulcorner \text{might } p \urcorner$ in c_1 and the proposition expressed by $\ulcorner \text{not-might } p \urcorner$ (its sentential negation) in c_2 are contradictories (and hence at least one must be false). Thus, the strategy of relaxing our assumption so far about deflationism about the truth predicate will not help either contextualism or relativism avoid the challenge raised by our data.¹¹

Having considered and responded to two pressing concerns about our methodology, we conclude that the data from this section presents a challenge to both contextualist and relativist theories. We

¹¹There is a version of relativism that would overcome this latter problem. The strategy would be to let $\ulcorner \text{might } p \urcorner$ be ambiguous between an assessment-sensitive reading and a non-assessment-sensitive reading (the latter would be similar to our KNOWER ORIENTED readings we discuss below in §4). This would then permit readings of $\ulcorner \text{might } p \urcorner$ and $\ulcorner \text{not-might } p \urcorner$ that are compatible. Despite the potential for empirical coverage, we think positing ambiguity here is less plausible than opting for genuine context sensitivity (as we do below), insofar as the pattern of data is robust across a range of linguistic expressions. Since we do not yet have data beyond epistemic modals, we will not press this point further at this time.

turn next to a different approach to thinking about modal semantics that we think is independently plausible and can account for the data discussed in this section.

4 Contextualist situation semantics

Contextualism is compatible with the results of our **Modal Utterances** case, since it allows for the possibility that contrary epistemic modal claims made in different contexts may be such that neither is false. However, in the **Modal Assessments** case, contextualism does not allow for contrary assessments to be such that neither is false, and this is because it predicts that an utterance of an epistemic modal sentence expresses a single proposition that has a unique truth value relative to each world. Relativism is not compatible with either the results of our **Modal Utterances** case or the **Modal Assessments** case, since it does not allow for the possibility that contrary epistemic modal claims (or contrary assessments of a single epistemic modal claim) made in different contexts may be such that neither is false.

So, neither contextualism nor relativism as traditionally formulated are in a position to predict the pattern of judgments of incompatibility that ordinary speakers in fact exhibit. In this section, we reflect on what kind of theory *could* predict these judgments. The most straightforward strategy for predicting our data would be a theory that is contextualist twice over: allowing (i) that the same epistemic modal sentence may express different propositions in different contexts, and (ii) that truth/falsity-assessments targeting the same object of assessment may express different propositions in different contexts. Allowing (i), the theory would be in a position to allow that $\ulcorner \text{might } p \urcorner$ and its negation $\ulcorner \text{not-might } p \urcorner$ express incompatible propositions when uttered (respectively) in contexts c and c' , but express compatible propositions when uttered (respectively) in contexts c'' and c''' . A theory allowing for these possibilities would be in a position to predict our data from **Two Utterances** case. Allowing (ii), the theory would be in a position to allow that, when targeting the same utterance, $\ulcorner \text{What } X \text{ said is true} \urcorner$ and $\ulcorner \text{What } X \text{ said is false} \urcorner$ express incompatible propositions when uttered (respectively) in contexts c and c' , but express compatible propositions when uttered (respectively) in contexts c'' and c''' . A theory allowing for these possibilities would be in a position to predict our data in the **Two Assessments** case.

In the rest of this section, we sketch a theory of this form. We want to emphasize that our aim here is to illustrate one way of predicting our the data, *not* to argue that our theory is the only or even best way to do so. Thus, we ask that the reader take our proposal here as a reasonable first pass attempt to make sense of and formally account for the pattern of judgments in §3.

4.1 The structure of the theory

Our theory departs from the simple versions of contextualism and relativism above in two ways. First, it holds that the inputs to domain-fixing functions are more fine-grained than possible worlds, and that domain-fixing functions are sensitive to differences between these more fine-grained entities. We will use situations (letting ‘*s*’ be a variable over situations) for this purpose, although we could have equally well used centered worlds (as in Egan (2007)). We are thinking of situations as possibly partial worlds—that is, each situation determines a unique world, but there may be distinct situations which determine the same possible world. Letting our semantics be sensitive to situations rather than worlds, we replace SIMPLE SEMANTICS with:

SITUATION SEMANTICS:

$$\llbracket \ulcorner \text{might } p \urcorner \rrbracket^{f,s} = 1 \text{ iff } \exists w \in f(s): \llbracket p \rrbracket^{f,w} = 1$$

On this account, then, the semantic values of declarative sentences are true relative to situations rather than worlds. There are various ways to generate a semantic content for $\ulcorner \text{might } p \urcorner$ out of this semantic theory. For instance, we could let context initialize the domain-fixing function, and let the semantic content of a declarative sentence be a set of situations:

C-CONTENT:

$$\text{The content of } \ulcorner \text{might } p \urcorner \text{ as uttered in } c \text{ is } \{s : \llbracket \ulcorner \text{might } p \urcorner \rrbracket^{c,f_c,s} = 1\}.$$

Or, we could invert this operation, letting context initialize the relevant situation and let the semantic content be a set of domain-fixing functions:

E-CONTENT:

$$\text{The content of } \ulcorner \text{might } p \urcorner \text{ as uttered in } c \text{ is } \{f : \llbracket \ulcorner \text{might } p \urcorner \rrbracket^{c,f,s_c} = 1\}.$$

Finally, we could let context initialize neither, and let the semantic content be a set of pairs of domain-fixing functions and situations:

R-CONTENT:

The content of \ulcorner might $p\urcorner$ as uttered in c is $\{\langle s, f \rangle : \llbracket \ulcorner$ might $p\urcorner \rrbracket^{c,f,s} = 1\}$.

The labels for these views were not chosen randomly. Roughly, and given the above situation semantic framework, C-CONTENT corresponds to the content a contextualist would assign \ulcorner might $p\urcorner$, E-CONTENT is similar to the content an expressivist would assign \ulcorner might $p\urcorner$ (see Yalcin (2007): 1010-1011), and R-CONTENT corresponds to the content a relativist (or non-indexical contextualist) would assign \ulcorner might $p\urcorner$ (see MacFarlane (2011, 2014)).

Notice immediately that R-CONTENT will not be able to predict the content of \ulcorner might $p\urcorner$ to vary across contexts of utterance, and hence a relativist situation semantics will be unable to predict the ambivalence in incompatibility judgments across conflicting utterances (at different contexts). However, we admit that, in principle at least, either C-CONTENT or E-CONTENT could work. We do not have grounds to make a principled decision between these two implementations of SITUATION SEMANTICS at this time. Nonetheless, for the sake of concreteness, we will opt for C-CONTENT in what follows. We again want to emphasize that this choice does not reflect a principled decision—we make it merely because this version offers a straightforward way to understand the contents of epistemic modal claims: according to it, they are sets of situations (roughly equivalent to Egan (2007)’s proposal that the contents of epistemic modal claims are sets of centered worlds).¹²

With a particular implementation of SITUATION SEMANTICS in hand, we come to the second component of the theory. Here, we reject our assumption of deflationism about the truth/falsity predicate in favor of a contextualist theory of truth. We propose that, as used in a context c , “true”/“false” predicates of a proposition that it is true/false at the situation of that context, s_c . Thus, given our assumption that “What Z said” refers to the proposition Z asserted, we have:¹³

¹²While we think our empirical evidence does not decide between these implementations, other factors might. For instance, theoretical considerations about content might be relevant (cf. Yalcin (2012)). Similarly, considerations from modal cognition may be relevant (cf. Phillips & Knobe (Forthcoming)). We set aside these matters for now.

¹³One may have noticed that CONTEXTUALIST TRUTH will predict that truth-/falsity-ascriptions are either necessarily true or necessarily false. One way to avoid this problem and still hold that truth-/falsity-ascriptions convey information about particular situations is to appeal to counterpart relations among situations. Thus, instead of in-

CONTEXTUALIST TRUTH

$$\llbracket \text{“What Z said is true”} \rrbracket^c = \{s: \llbracket \text{“What Z said”} \rrbracket^{c, f_c, s_c} \text{ is true at } s_c\}.$$

We turn next to show that C-CONTENT + CONTEXTUALIST TRUTH can (in principle) predict the pattern of data we found in §3. Remember that the data we aim to predict here is the possibility of agreement and disagreement with (Q) for both the **Modal Utterances** and **Modal Assessments** cases (allowing us to capture the observed midpoint agreement in these conditions):

(Q) At least one of the inspectors’ claims must be false.

Start with **Modal Utterances** case, in which Inspector A says (7) and Inspector B says (8), in different contexts.

(7) Fat Tony could have died at the docks.

(8) Fat Tony couldn’t have died at the docks.

Since it is a version of contextualism, our theory allows that A and B’s claims here may, or may not, express compatible propositions. If $f_{c_A} = f_{c_B}$, then A and B’s claims will be incompatible.¹⁴ But if $f_{c_A} \neq f_{c_B}$, then A and B’s claims will not be incompatible. Hence, our contextualist situation semantics is compatible with agreement with (Q) and disagreement with (Q).

Turn next to the **Modal Assessments** case, in which the Chief utters (7), and Inspectors A and B (in different contexts from the Chief and each other), overhear the Chief’s utterance and say, respectively:

(9) What the Chief said is true.

(10) What the Chief said is false.

dexing “true” and “false” to a particular situation, we rather index them to a function from situations to situations, g_s , that is anchored to a particular situation, s . The idea is that g_s maps a situation s' to the counterpart of s that is world-mates with s' . Then, we get that $\llbracket \text{“}\phi \text{ is true”} \rrbracket^c = \{s: \llbracket \phi \rrbracket^{c, f_c, s_c} \text{ is true at } g_{s_c}(s)\}$. *This* proposition will be nontrivially true/false, owing to the fact that for different situations s', s'' , $g_{s_c}(s') \neq g_{s_c}(s'')$.

¹⁴Note that c_A is not the context of assessment c_A . We distinguish these by using non-italic font for Inspector A’s context.

In this case, our theory delivers a single proposition expressed by the Chief: that it is compatible with the information in the Chief’s context that Fat Tony died at the docks ($= \{s : \exists w' \in f_{c_C}(s) : \llbracket \text{“Fat Tony died at the docks”} \rrbracket^{c_C, f_{c_C}, w'} = 1\}$). Whether our theory predicts that (Q) is true or (Q) is false depends on the nature of f_{c_C} .

The first kind of possibility is that f_{c_C} does *not* vary in output across situations at the same world. In that case, our theory will predict that A and B’s utterances are contradictory and so one must be false. That is because, since A and B’s utterances take place at the same world, no matter what situations they target, $s, s', f_{c_C}(s) = f_{c_C}(s')$ and so one of the claims must be false.

The second kind of possibility is that f_{c_C} *does* vary in output across situations at the same world. Then, since A and B’s assessments target different situations (saying of the Chief’s claim that it is true at one and false at another, respectively), in principle, both of their claims may be true. A’s claim is that $f_{c_C}(s_A)$ is compatible with Fat Tony having died at the docks, while B’s claim is that $f_{c_C}(s_B)$ is not compatible with Fat Tony having died at the docks. Both of these claims may be true, as long as $f_{c_C}(s_A) \neq f_{c_C}(s_B)$. Therefore, SITUATION SEMANTICS + C-CONTENT + CONTEXTUALIST TRUTH allows that, in the **Modal Assessments** case, neither what Inspector A says nor what Inspector B says must be false.¹⁵

4.2 Fleshing out the theory

We have seen how neither a standard contextualist nor relativist theory is in a position to predict the incompatibility intuitions of ordinary speakers regarding epistemic modal claims. But we have

¹⁵Again, we want to emphasize that our claim is not that this kind of theory is the *only* way to predict the data from §3. Here are two alternative strategies that one could begin to pursue. The first would be to adopt an expressivist line, thus holding that the claims made by uttering epistemic modal sentences are (strictly speaking) neither true nor false. Ordinary speakers implicitly recognize this, so when confronted with the question whether at least one of the claims must be false, they charitably look for an alternative interpretation of what is being asked, and land on two different possible interpretations, either interpreting the question as asking about whether at least one must be false relative to their (the reader’s) context, or interpreting the question as asking about whether at least one must be false relative to their (the utterers) respective contexts. The former question would then be correctly answered in the affirmative and the latter correctly answered in the negative. This strategy is similar to Yalcin (2011)’s discussion of ambivalence in eavesdropping scenarios. It remains to be seen whether the strategy could be expanded upon to explain our incompatibility results, and this is an issue we hope to explore in future work. A second alternative strategy could follow the “cloudy contextualist” theory of von Steup & Gillies (2011). According to that theory, when someone utters \lceil might p \rceil they “put in play” a set of propositions, each resolving the contextual indeterminacy as to the domain-fixing parameter differently. This theory holds some promise in explaining our results, as long as sense can be made of what it would be for two modal claims (or assessments thereof) to be incompatible or compatible on that view.

also seen that a contextualist situation semantics, together with a contextualist theory of truth, can. Structurally, such a theory allows for two dimensions of variability in the truth conditions of epistemic modal claims: the situation of evaluation, and the modal's domain fixing function. But a question immediately arises: what is the intuitive picture of epistemic modal discourse, on such a theory?

For illustrative purposes, consider again the **Modal Utterances** case from our experiment, where Inspector A looks at the evidence at the docks and concludes that Fat Tony could have died at the docks, while Inspector B sees Fat Tony plant the evidence and concludes that Fat Tony could not have died at the docks. When discussing their two claims, a natural thing to think is that they are really making two different claims. Knowing only about the evidence at the docks, Inspector A made a claim about what was possible given what he knows from the evidence at the docks. Knowing only about the evidence from the security camera, Inspector B made a claim about what was possible given what he knows. There's a natural sense in which these are just two different claims, and it's not hard to imagine a continuation of the story that (tacitly) concerns exactly this difference. For example, the Chief might commend both Inspectors on their conclusions: Inspector A because he correctly assessed which possibilities had been left open by what he learned at the docks; Inspector B because he correctly assessed which possibilities had been left open by what he learned from the security camera. On this intuitive picture, there is no reason that one of the Inspectors *must* have been wrong, since they were making two different claims, and their claims concerned different sets of possibilities.

Next, turn to the **Modal Assessments** case. Here, too, there is a way for neither Inspectors' *assessments* of the Chief's claim to be false. Suppose the Chief's claim that Fat Tony could have died at the docks is not about any particular body of evidence, but rather varies in truth value depending on what evidence is available in a given situation. Inspector A states that the Chief's claim is true; Inspector B states that the Chief's claim is false. Importantly, their two situations differ in an important way, since Inspector B's situation involves evidence that rules out Fat Tony dying at the docks, but Inspector A's situation does not. Thus, it is plausible that their truth/falsity-ascriptions are tacitly about their respective situations. Once again, it is not

difficult to imagine a conversation that (tacitly) admits to this difference. Suppose, for example, that the Chief has learned about both of the Inspectors’ assessments of his statement on the news, and, enjoying a good fight, the Chief tells the two inspectors that they contradicted each other. It would be natural for Inspector A to respond by pointing out that his claim wasn’t at all meant to contradict Inspector B’s. After all, in the situation he was in at the time, there was simply no evidence that Fat Tony had planted the evidence. Only Inspector B knew that. The situations they were in when evaluating the Chief’s claim differed in such a way that they weren’t actually contradicting each other.

We’ve tried to sketch an intuitive picture of what the two dimensions are and how they may play a role in ordinary modal discourse. So, how does this intuitive picture line up with our formal semantics from the last section? One approach holds that there are broadly two kinds of epistemic domain fixing functions: *situation-variant* and *knower-oriented*. There is one situation-variant function, call it f^* , which is characterized as follows:

SITUATION VARIANT

For any s , $f^*(s) = \{w : w \text{ is compatible with the evidence available in } s\}$.

The situation variant modal base f^* maps situations to the worlds compatible with the available evidence in those situations.¹⁶ Compare the situation-variant modal base with a knower-oriented one, defined as follows:

KNOWER ORIENTED

For some epistemic agent N , time t , and situation s : $f_{N_t}(s) = \{w : w \text{ is compatible with what } N \text{ knows at } t \text{ at } w_s\}$.

Knower-oriented epistemic domain fixing functions are relativized to epistemic agents at particular times, and characterize the informational content of what such agents know (or have as evidence) at various worlds at that time. Basically, situation-variant and knower-oriented domain functions differentially separate the role of the inquirer in determining some set of epistemically possible

¹⁶There is an immediate issue of the content of “available” here. We do not want to get into these details at this time. The more general theory will likely involve different situation-variant modal bases, each utilizing a different notion of availability. See Hacking (1967); DeRose (1991); Dowell (2013).

worlds. The situation-variant domain function lets what is epistemically possible vary from situation to situation, and thus vary from inquirer to inquirer. By contrast, the knower-oriented domain function makes what is epistemically possible depend entirely on what is known to some particular agent at some particular time.¹⁷

We can now combine this hypothesis about modal domain fixing functions with our contextualist situations semantics and contextualist semantics for “is true”/“is false” to predict the pattern of data we found in §3. Basically, in each vignette, there will be two kinds of interpretive possibilities, and depending on how people opt for those interpretive possibilities, this is what will generate the midpoint level agreement with the incompatibility judgments (Q).

Take the **Modal Utterances** case first. Either Inspector A and B’s contexts initialize the same domain-fixing function, or different domain-fixing functions. If they initialize the same domain-fixing function, then for any situation s : one of their claims is false at s . If they initialize different domain-fixing functions, then for some situations s it may be that neither’s claim is false at s .¹⁸ Next, turn to the **Modal Assessments** case. The Chief’s context initializes a particular modal base, either the situation variant f^* or the knower oriented f_{C_t} . Inspectors A and B are making truth-claims targeting the Chief’s assertion, and their claims are about some situation. If the Inspectors are making claims about the same situation (most likely the Chief’s), then at least one

¹⁷One might wonder whether there is independent evidence for positing two kinds of modal bases, that is, if there is independent evidence for distinguishing an interpretation of a modal sentence on which its truth value varies across situations (our SITUATION VARIANT interpretation) from an interpretation on which its truth value does not vary across situations (our KNOWER ORIENTED interpretation). We think the more nuanced data on retraction discussed in von Fintel & Gillies (2011) might provide such evidence. von Fintel & Gillies (2011) point out that, when confronted with evidence that p is false, it seems possible to react by retracting your claim, but also possible to stick to ones guns:

A: The keys might be in the drawer.

B: No, I have them with me. Why did you say that?

A1: Oh, I guess I was wrong.

A2: I didn’t say they *were* in the drawer, only that they *might* be, and they might have been!

Here, (A1)’s response indicates that she had intended the situation-variant interpretation of her modal claim: one whose truth value changes as she learns more information. By contrast, (A2)’s response indicates that she had intended the knower-oriented interpretation of her modal claim, and that it was centered on the information she had when uttering it.

¹⁸There are three ways f_{c_A} may be different from f_{c_B} : (i) $f_{c_A} = f^*$ and $f_{c_B} = f_{B_t}$ (the function that yields the set of worlds compatible with what B knows at the time of utterance t at w_s , for any s), (ii) $f_{c_B} = f^*$ and $f_{c_A} = f_{A_t}$ (the function that yields the set of worlds compatible with what A knows at the time of utterance t at w_s , for any s), and (iii) $f_{c_A} = f_{A_t}$ and $f_{c_B} = f_{B_t}$.

of their claims is false. But if the Inspectors are making claims about different situations *and* the Chief’s modal base is f^* , then both of their claims may be true (A’s claim then is that it is compatible with A’s evidence that Fat Tony died at the docks, while B’s claim that it is not compatible with B’s evidence that Fat Tony died at the docks).

5 Conclusion

The aim of our paper is to motivate and implement a new way of assessing various semantic/pragmatic theories of epistemic modals. We first argued that the current focus on eavesdropper truth-value intuitions is misguided because both contextualist and relativist theories are compatible with the data, no matter how it turns out. Then, we motivated an alternative way to evaluate contextualism and relativism by exploring inter-contextual compatibility judgments. Implementing this strategy, we found a pattern of ordinary judgments that is not predicted by standard contextualist and relativist theories. We ended with a novel contextualist situation semantics which *does* predict the pattern of judgments we found.

Stepping back from the particulars of the contextualism/relativism debate over the semantics of epistemic modals, it bears pointing out that the phenomenon we have uncovered is likely to extend beyond epistemic modal expressions. A much wider range of expressions are known to exhibit some kind of contextual variability (normative expressions, predicates of taste, gradable adjectives, quantifiers, conditionals, and so on), but don’t obviously pattern with so-called “automatic indexicals”, like first person pronouns. A promising possibility is that the present proposal can be fruitfully be extended to this wider class of expressions, which have been of central interest across philosophy, linguistics and cognitive science. Critically, the methodology we developed in this paper should be straightforwardly applicable to this wider class of expressions, allowing us to empirically test whether these expressions exhibit the same kind of contextual variation that we have observed in epistemic modals, and to semantically account for this variation where it arises. This promising extension of the present proposal should not only provide a clearer empirical picture of the context-sensitivity of these expressions themselves, but would shed light on the extent and scope of the underlying phenomenon that gives rise to this kind of pattern in our use of epistemic

modals.

References

- DeRose, Keith. 1991. Epistemic Possibilities. *The Philosophical Review*, **100**(4), 581–605.
- Dietz, Richard. 2008. Epistemic Modals and Correct Disagreement. *Pages 239–62 of: Garcia-Carpintero, Manuel, & Kölbel, Max (eds), Relative Truth*. Oxford: Oxford University Press.
- Dowell, Janice. 2011. A Flexible Contextualist Account of Epistemic Modals. *Philosophers' Imprint*, **11**(14), 1–25.
- Dowell, Janice. 2013. Flexible Contextualism about Deontic Modals: A Puzzle about Information-Sensitivity. *Inquiry*, **56**(2-3), 149–178.
- Egan, Andy. 2007. Epistemic Modals, Relativism, and Assertion. *Philosophical Studies*, **133**(1), 1–22.
- Egan, Andy. 2010. Relativism about Epistemic Modals. *In: Hales, Steven (ed), Blackwell Companion to Relativism*. Oxford: Blackwell.
- Egan, Andy, Hawthorne, John, & Weatherson, Brian. 2005. Epistemic Modals in Context. *Pages 131–170 of: Preyer, George, & Peter, George (eds), Contextualism in Philosophy: Knowledge, Meaning, and Truth*. Oxford: Oxford University Press.
- von Fintel, Kai, & Gillies, Anthony. 2011. Might Made Right. *Pages 108–130 of: Egan, Andy, & Weatherson, Brian (eds), Epistemic Modality*. Oxford: Oxford University Press.
- Hacking, Ian. 1967. Possibility. *The Philosophical Review*, **76**, 143–68.
- Katz, Jonah, & Salerno, Joe. 2016. Epistemic Modal Disagreement. *Topoi*, 1–13.
- Khoo, Justin. 2015. Modal Disagreements. *Inquiry*, **58**(5), 511–534.
- Khoo, Justin, & Knobe, Joshua. 2016. Moral Disagreement and Moral Semantics. *Noûs*.
- Knobe, Joshua, & Yalcin, Seth. 2014. Context-sensitivity of epistemic possibility modals: experimental data. *Semantics & Pragmatics*, **7**(4), 1–21.
- Kratzer, Angelika. 1981. The Notional Category of Modality. *Pages 38–74 of: Eikmeyer, H. J., & Rieser, H. (eds), Words, Worlds, and Contexts. New Approaches in Words Semantics*. Berlin: de Gruyter.
- Kratzer, Angelika. 1991. Modality. *Chap. 23, pages 639–650 of: von Stechow, Arnim, & Wunderlich, Dieter (eds), Handbuch Semantik*. Berlin and New York: de Gruyter.
- MacFarlane, John. 2009. Nonindexical Contextualism. *Synthese*, **166**(2), 231–250.
- MacFarlane, John. 2011. Epistemic Modals are Assessment-Sensitive. *Pages 144–178 of: Egan, Andy, & Weatherson, Brian (eds), Epistemic Modality*. Oxford: Oxford University Press.

- MacFarlane, John. 2014. *Assessment Sensitivity: Relative Truth and its Applications*. Oxford: Oxford University Press.
- Phillips, Jonathan, & Knobe, Joshua. Forthcoming. The Psychological Representation of Modality. *Mind & Language*.
- Sarkissian, Hagop, Park, John, Tien, David, Wright, Jennifer Cole, & Knobe, Joshua. 2011. Folk Moral Relativism. *Mind & Language*, **26**(4), 482–505.
- Stanley, Jason. 2005. Fallibilism and Concessive Knowledge Attributions. *Analysis*, **65**(2), 126–131.
- Stephenson, Tamina. 2007. Judge Dependence, Epistemic Modals, and Predicates of Personal Taste. *Linguistics and Philosophy*, **30**(4), 487–525.
- Yalcin, Seth. 2007. Epistemic Modals. *Mind*, **116**(464), 983–1026.
- Yalcin, Seth. 2011. Nonfactualism About Epistemic Modality. *Pages 295–332 of: Egan, Andy, & Weatherson, Brian (eds), Epistemic Modals*. Oxford: Oxford University Press.
- Yalcin, Seth. 2012. Bayesian Expressivism. *Pages 123–160 of: Proceedings of the Aristotelian Society*, vol. 133.