2022. *Proc Ling Soc Amer* 7(1). 5249. https://doi.org/10.3765/plsa.v7i1.5249.

Accounting for variability in the truth-evaluation of bare epistemic possibility statement

Giuseppe Ricciardi & Joshua Martin*

Abstract. It is hotly debated whose perspective is relevant for defining the truth-value of bare epistemic possibility statements: the utterer or the assessor. Central to this debate are findings on truth-value judgments of 'might p' statements in "eavesdropping" scenarios where the statement is appropriately asserted from the point of view of the speaker but does not correspond to how in reality things are. We offer findings from two studies suggesting that in these scenarios English speakers disagree on the truth-value not only of 'might p' but also, surprisingly, of bare 'p'. We argue that underlying this behavior is not a disagreement on the semantic content of the statements but a disagreement on how to interpret the meaning of the adjectives 'true' and 'false' when applied to a statement uttered by another agent. Specifically, we compare two potential accounts: the first account ascribes to the adjective 'true' an ambiguity between a 'coherent' sense – a statement uttered by A is true as long as the statement coheres with A's evidence – and a 'correspondentist' sense - a statement as uttered by A is true as long as the statement correspond to reality; the second account treats 'true' as having only the correspondentist sense but ascribes to it an inherent relative nature – to decide if a statement as uttered by A is true one has to first decide a perspective (A or the assessor). Lastly, we discuss the implications of the findings for the semantics of epistemic 'might p' statements.

Keywords. epistemic modality; possibility; truth; experimental semantics

1. Introduction. Consider the dialogue in (1).

(1) Context: Susan, John, and Mark are friends. They are all college students. It's 5pm on Wednesday. Susan and Josh meet in the university gym.

John: "I'm surprised to not see Mark: usually he is here at this time on Wednesday"
Susan: "**He is working in the library**"

Imagine that you overheard this dialogue and you happen to know that Mark is not working in the library now, but he is home in his bed with a fever. In this situation, if you were asked to judge whether the statement made by Susan is true or false, what would you answer? For anyone trained in formal semantics the answer should be straightforward: 'false'. This judgment is based on the assumptions that what makes a statement 'true' or 'false' is the correspondence between the 'proposition' conveyed by the statement and how things are (in this case things are such that Mark is not working in the library at the time of Susan's utterance). Let's call this the standard use of 'true statement' in formal semantics.

Now, imagine that you ask Susan why she said that Mark is working in the library, and she tells you that she had met Mark earlier in the morning and he had told her that today from 3 to 8pm he would start his part-time job at the library. Mark didn't lie to her but, unbeknown to Susan, he got sick over lunch and could not make his shift. After you know the information based

* We would like to thank for their feedback on this project Kate Davidson, Deniz Rudin, Alexis Wellwood, Shannon Bryant and the audience of Harvard Meaning & Modality Lab and The Meaning Lab at USC. Authors: Giuseppe Ricciardi, Harvard University (gricciardi@g.harvard.edu) & Joshua Martin, Harvard University (joshuamartin@g.harvard.edu).

on which Susan made her statement, would you consider that Susan was justified in making her statement? If yes, would you still judge the statement as 'false'? According to the standard formal semantic use of 'true statement', you should still judge the statement false because the proposition doesn't make any reference to the speaker's evidence supporting the statement: even though Susan was justified in making her statement, yet the proposition conveyed by the statement doesn't correspond to how things are.

Let's imagine, now, that instead of making a bare 'p' statement Sarah said a statement containing 'might' in its epistemic use as in (2).

(2) Context: same as in 1.

John: "I'm surprised to not see Mark: usually he is here at this time on Wednesday"
Susan: "**He might be working in the library**"

When epistemic 'might' (or any other epistemic expression) occurs in a statement, is the speaker's evidence supporting the statement part of the conveyed proposition? An influential line of thought - which can be traced back to Kratzer (1977) - answers positively to this question. More specifically, according to this view, Sarah's statement, which appears to be 'bare', includes a non-overtly spelled-out component that refers to the body of evidence K relative to which Sarah is making her modal judgment: "[Relative to the body of evidence K] He might be working in the library". Within the possible world framework adopted by Kratzer, this hypothesis consists of defining the proposition expressed by Sarah's statement roughly as: 'there is at least a possible world compatible with the body of evidence K such that John is working in the library in that world at the time of utterance'. More generally, in this view, any utterance of a bare epistemic 'MODAL p' statement expresses a proposition of the type 'Relative to (from the perspective of) the body of knowledge K, MODAL that p'. This view is usually called 'the standard contextualist approach to the semantics of epistemic modality', where 'contextualist' is meant to indicate that the relevant body of evidence for interpreting the modal is established in the context of utterance (e.g., Kratzer 1977, 1981, 1991, 2012).

This standard contextualist hypothesis makes the prediction that if one were asked, for example, to assess Susan's statement in (2) as 'true' or 'false', one would first try to find out what exactly K is in this case and then check whether based on K it is possible that John is working at the library now. Assuming that for a matrix bare epistemic statement K is virtually always some evidence available to the speaker, then this hypothesis makes the prediction that in order to assess the truth of a matrix bare epistemic statement like in (2) it does matter which evidence the speaker was relying on making the statement because that evidence is part of the conveyed proposition. Thus, the Kratzerian view predicts an asymmetry between the truth-evaluation of bare 'p' statements and bare epistemic 'might p' statements: while in the bare 'p' case the speaker's evidence doesn't matter, in the epistemic case it does matter.

This prediction has been explicitly challenged: "people tend to assess epistemic modal claims for truth in light of what they (the assessors) know, even if they realize that they know more than the speaker (or relevant group) did at the time of utterance" (MacFarlane 2011: 20). Although this observation is phrased as applying generally to epistemic statements, the literature exemplifies the claim only with bare epistemic 'might p' statements. Here, we follow this practice of focusing on 'might p'. Based on this intuition some scholars have advanced a hypothesis which is known as 'relativist approach to the semantics of epistemic modality', where 'relativist' refers to the fact that the truth-value of an epistemic 'might p' statement is relative to the perspective adopted by whoever is performing the assessment (e.g., MacFarlane 2011, 2014; Egan

2007; Egan et al. 2007; Stephenson 2007). The current consensus among scholars of epistemic modality is that this empirical observation is right; however, not everyone is of the opinion that this data point compels one to abandon the contextualist hypothesis altogether: some scholars, instead, suggest that what needs to be abandoned is just the standard version of the contextualist view (e.g., Yalcin 2007, 2011; von Fintel & Gillies 2008, 2011; Dowell 2011; Yanovich 2014; Stalnaker 2014; Ninan 2018) or the standard view of assertion of a statement (e.g., Rudin 2021). Thus, the current consensus is that there is a symmetry between the truth-evaluation of bare 'p' statements and of bare epistemic 'might p' statements as uttered by an agent A: in both cases A's evidence supporting the statement doesn't matter.

In this work, we aim at assessing this very empirical generalization by asking the following question: when English speakers are asked to evaluate the truth of a 'p' or a 'might p' statement as uttered by an agent A, do they consider in their evaluation A's body of information at the time of utterance? To answer this question, we report here on findings from two experiments which adopted a version of the "eavesdropping task" previously employed in Knobe & Yalcin (2014), Phillips & Khoo (2019), Phillips & Mandelkern (2020), and Reuter & Brun (2021). Specifically, we focus here on the task adopted in Knobe & Yalcin (2014)'s Experiment 4: first participants read a vignette where a character utters a justified statement - either 'p' or 'might p' - but then they are told that in reality 'not-p'; next, after reading this story, they were asked to judge whether the statement is true or false. If participants don't consider the speaker's evidence supporting a statement - either 'p' or 'might p' - when evaluating its truth-value, then we would expect that in both cases they would uniformly judge the statement 'false' given that 'not-p'. But Knobe & Yalcin (2014) found that while in the 'p' case participants uniformly judged the statement 'false', in the 'might p' case they didn't converge towards a clear response. However, the findings about the 'p' conditions are at odds with what reported in Reuter & Brun (2021): they asked participants to truth-evaluate a 'p' statement in a similar eavesdropping task and found that even in this case participants didn't converge towards a clear response.

Here, we offer two follow-up studies to Knobe & Yalcin (2014) aimed at assessing whether their findings were due to confounding factors: in Experiment 1 we assess the 'might p' case and in Experiment 2 we assess the 'p' case. To preview, our findings from Experiment 1 replicate with a modified task and scenario the findings in Knobe & Yalcin (2014) about 'might p'; our findings from Experiment 2 show that participants' convergence towards 'false' in Knobe & Yalcin's (2014) 'p' condition was due to the fact that participants judged the character not justified to make the statement; instead, once one makes sure that participants judge the character justified in making the statement, it turns out that a significant (although still a minority) portion of participants shifts towards 'true', making the behavior in the 'p' and the 'might p' case look much more alike, as already suggested by the findings in Reuter & Brun (2021). So, before considering the 'might p' case, what needs to be explained is the surprising (from the point of view of formal semantics) behavior in the 'p' case: a significant portion of English speakers are willing to judge as 'true' a 'p' statement not corresponding to how things are but justified from the point of view of the character uttering the statement. We argue that underlying this behavior is not a disagreement on the semantic content of the statements but a disagreement on how to interpret the meaning of the adjectives 'true' and 'false' when applied to a statement uttered by another agent. Specifically, we compare two potential accounts: the first account ascribes to the adjective 'true' an ambiguity between a 'coherent' sense – a statement uttered by A is true as long as the statement coheres with A's evidence – and a 'correspondentist' sense - a statement as uttered by A is true as long as the statement correspond to reality; the second account treats

'true' as having only the correspondentist sense but ascribes to it an inherent relative nature – to decide if a statement as uttered by A is true one has to first decide a perspective (A or the assessor). Lastly, we discuss the implications of the findings for the semantics of epistemic 'might p' statements.

The paper is organized as follows. In section 2, we review previous experimental studies that have addressed this very question. In section 3 we offer novel experimental findings. In section 4, we offer a general discussion of our findings and of their relevance for the discussion about epistemic possibility statements. In section 5, we offer some concluding remarks.

2. Previous studies. Knobe & Yalcin (2014) designed a battery of experiments using the eavesdropping task. Their Experiment 4 can be considered the most critical in that it used a scenario that has been prominent in the literature and its findings were recently replicated in Phillips & Mandelkern (2020). So, we focus on this experiment. The task prompt participants to read a story where a character has some reasons to utter a statement - either 'might p' ('Modal' condition) or 'p' (Nonmodal condition) - and right after we are told that actually 'not-p' (see 3).

(3) *Vignette in Knobe & Yalcin's (2014) Experiment 4*
Sally and George are talking about whether Joe is in Boston. Sally carefully considers all the information she has available and concludes that there is no way to know for sure. Sally says: "Joe might be in Boston." [Modal] / "Joe is in Boston" [Nonmodal]
Just then, George gets an email from Joe. The email says that Joe is in Berkeley. So, George says: "No, he isn't in Boston. He is in Berkeley."

In one of the conditions ("Falsity") participants were asked to perform a truth-value judgment of Sally's utterance. The question was asked in this form: "We want to know whether what Sally said is false. So please tell us whether you agree or disagree with the following statement: 'What Sally said is false'". Participants were asked to indicate their level of agreement on a scale from 1 ("completely disagree") to 7 ("completely agree"). Knobe & Yalcin (2014) found that while whereas the mean agreement rating for the nonmodal statement was around 6, the mean agreement rating for the modal statement was around 3. The authors interpret these two findings as suggesting that, while there is convergence among participants in judging Sally's statement false when the statement is 'p', there is a lack of agreement among participants as to whether Sally made a false statement when the statement is 'might p'. So, based on these findings, one can be tempted to conclude that, at least for the bare 'p' statements, adult English speakers don't take in consideration the speaker's perspective at the time of utterance when invited to evaluate the truth of the statement. However, this conclusion is too hasty. In fact, Reuter & Brun (2021) report findings suggesting a different picture about the behavior of English speakers when truth-evaluating bare 'p' statements in an eavesdropping task.

In their experiment 1, Reuter & Brun (2021) designed two vignettes by adopting the following rationale: in both cases the story features a character making a bare objective 'p' statement in a situation where the information available to the character at the time of utterance is good enough to license the statement but, at the same time, the event described by 'p' does not hold. Here are the two stories.

(4) *The two stories used in Reuter & Brun's (2021) Experiment 1*

[Party] Anne and Robert go to a party late at night. On their way to the party, Anne asks Robert whether any of his friends are at the party. Robert answers that Jill is at the party,

because Jill had told Robert a few hours before that she would go. When they arrive at the party, it turns out that Jill had changed her plans, and actually is not at the party"

[Rolex] Maria is a watch collector. She keeps all her watches in a safe and knows her collection really well. One day, her friend John asks her whether she has a 1990 Rolex Submariner in her safe and, if so, could show it to her. Maria answers that she has got a 1990 Rolex Submariner in her safe. After all, she had purchased that watch a few years ago. When Maria opens the safe a little later, she finds out that a burglar has stolen several watches, among them the 1990 Rolex Submariner

After reading either of these vignettes, each participant was asked one of the following two questions depending on the vignette they read: [Party] Was Robert's answer true or false? [Rolex] Was Maria's answer true or false? Participants were presented with three options: (1) true; (2) false; and (3) not sure. Surprisingly, in both scenarios most participants selected the option 'true': 59.6% in the Party case, and 56.8% in the Rolex case (as for the other two options: in the Party case 38.1% chose 'false' and 2.1% chose 'not sure; in the Rolex case 27.3% chose 'false' and 15.9% chose 'not sure'). In one of the conditions of their experiment 3, Reuter & Brun (2021) re-run the task focusing only on the Rolex scenario with these modifications: they gave participants only two response options 'true' and 'false' and had the truth question preceded by the question 'Did Maria answer the question to the best of her knowledge?''. The latter manipulation was based on the hypothesis that some participants might have answered 'true' in their first experiment because instead of the truth-question they answered a question about the assertability of the statement by the protagonist of the story: "when we first ask participants whether the protagonist of the scenario answered the question to the best of her knowledge, it is unlikely that the participants will still substitute the subsequent truth question with the same question they just answered" (Reuter & Brun 2021: 13-14). Interestingly, they found that, even though all participants answered 'yes' to the question whether Maria answered to the best of her knowledge, they still split in their answer to the question 'Was Maria's answer true or false?': 51.1% answered 'true' and 48.9% answered 'false'. At the least, these findings suggest that for some English speakers a bare objective 'p' statement is true as soon as its speaker made it to the best of their knowledge (i.e. its speaker was entitled to make it based on the information available to them) no matter whether the state of affairs described by 'p' holds in reality. What's underlying this behavior? We discuss the answer to this question in the general discussion in section 4.

To summarize, the findings in Knobe & Yalcin's (2014) Experiment 4 - replicated in Phillips and Mandelkern (2020) – suggest that English speakers uniformly disregard the speaker's evidence when truth-evaluating a 'p' statement in an eavesdropping task but are divided when it comes to truth-evaluating a 'might p' statement. However, the findings in Reuter & Brun's (2021) Experiment 1-3 suggest that English speakers are divided in truth-evaluating bare 'p' statements as well. In the next section, we offer two follow-up studies to Knobe & Yalcin's (2014) designed with the following goals: (i) to investigate whether the findings in their Modal condition were due to task effects (Experiment 1); (ii) to investigate why Knobe & Yalcin's (2014) findings in their Nonmodal condition differed from those in Reuter & Brun (2021) (Experiment 2).

3. Our studies.

3.1. EXPERIMENT 1. In this study, our goal was twofold: first, we wanted to investigate whether the findings in Knobe & Yalcin's (2014) Nonmodal-Falsity condition (i.e. truth-value judgment of 'might p') were due to task effects; second, we wanted to make the task more homogeneous

with the task in Reuter & Brun (2021) to facilitate a comparison between the two. To start, we identified two potential sources for the variability in truth-value judgment: one in the interpretation of the task and the other in the interpretation of the story. Concerning the interpretation of the task, notice that the original test question asked participants to express agreement with the target statement 'What Sally said is false'. We reasoned that, while some people may have interpreted the task as a truth-value judgment of the statement uttered by Sally (the task at stake in the theoretical discussion and the one intended by Knobe & Yalcin), some other people may have interpreted the task as an evaluation of the felicity of Sally's utterance. The hypothesis that non-linguistically trained speakers might not distinguish between truth-value judgments and felicity judgments has been advanced in recent literature concerned with methodological issues (e.g., Tonhauser & Matthewson 2015; Jasbi et al. 2019; Waldon & Degen 2020; Scontras & Pearl 2021). Under this interpretation, the variability found by Knobe & Yalcin (2014) could be due to participants adopting either interpretation: those who interpreted the task as a truth-value judgement assessed Sally's statement as false, whereas those who interpreted the task as a felicity judgement assessed Sally's statement as true. To test this hypothesis, we did two things.

First, we wanted to avoid using in the test question the anaphoric phrase 'what Sally said' which, in virtue of explicitly mentioning Sally, may prompt participants to focus on the felicity of Sally's communicative performance more than on the truth of what she said; so, we adopted from Doran et al. (2012) the strategy of presenting the story with the statement uttered by Sally as underlined and then refer back to it in the test question through the definite description 'the underlined statement'. With this move, the new test question would have looked like: 'We want to know whether the underlined statement is false. So please tell us whether you agree or disagree with the following statement: The underlined statement is false'. However, we judged the original way of posing the test question a little bit too convoluted in that it asked participants to indirectly judge the truth-value of a statement through agreeing with another statement which refers to the statement to be truth-evaluated. So, we decided to simplify the task by eliminating the intermediate step and directly asking 'Is the underlined statement true or false?'. We also offered only two response options ('True' or 'False') because we were interested in detecting which tendency would emerge when forcing participants to take either stance. Moreover, this way of asking the test question makes the task very similar to the one adopted in Reuter & Brun (2021).

Second, we also explicitly tested whether participants judge Sally's communicative act felicitous by asking "Based on what she knows, is Sally justified to say the underlined statement?" with the response options "Justified" and "Not Justified". We had participants see either judgement alone (Between-subjects condition) or both judgements with justification preceding truth (Within-subjects condition). The goal was to test whether there is convergence in judging Sally's communicative act felicitous independently from the truth-value judgement question and whether having participants first answering the felicity question had any effect on their performance in the truth-question. In this respect, we had a similar goal in mind as the one Reuter & Brun (2021) had in mind in designing one of the conditions in their experiment 3 when they asked participants 'Did Maria answer the question to the best of her knowledge?' before the truth question.

Concerning the interpretation of the scenario, we reasoned that in the original scenario designed in Knobe & Yalcin (2014) (see 3) there are three features which may be problematic: a) participants are not told any positive evidence for Sally's claim which makes Sally's utterance not completely natural in the first place; b) participants have to assume George's perspective which may increase the cognitive load; c) participants can decide to assess the statement based

on what they know at the utterance time T - when the assessor/participant is in the same epistemic position as the utterer and has no reason to judge the sentence false - or at the time T' - when participants learn that 'p' is false through the email received by the character George. We speculated that part of the variability could be due a split between participants assessing Sally's statement at T or a T'. So, we also manipulated between-subjects the type of conversation such that participants read either the original conversation or a variant of it with these changes: a) Sally's evidence supporting her 'might p' claim is made explicit; b) Sally is talking directly to the participant; c) the participant is told right away at the utterance time that 'p' is false. The full materials are reported in (5).

(5) Experiment 1: materials

Original story
Sally and George are talking about whether Joe is in Boston. Sally carefully considers all the information she has available and concludes that there is no way to know for sure. Sally says: 'Joe might be in Boston'. Just then, George gets an email from Joe. The email says that Joe is in Berkeley. So, George says: 'No, he isn't in Boston. He is in Berkeley.'

New story
Imagine that you are sitting in a café in Berkeley with a friend of yours named Sally, having a conversation about another friend, Joe. Sally says: "Joe might be in Boston, since he had a business meeting scheduled there this week". You know that Joe's meeting in Boston was canceled and he is in Berkeley right now.

Truth question: Is the underlined statement true or false?

Justification question: Based on what she knows, is Sally justified to say the underlined statement?

Putting together our manipulations, we obtained a 2X2X2 design with the following three factors each consisting of two levels: Story ('Original'; 'New') which was manipulated between-subjects; Test question ('Justification', 'Truth'); Number of test question per participant ('Between'; 'Within). We recruited 300 participants (200 for the four between conditions and 100 for the within conditions) on Mturk. We adopted the following criteria of exclusion: participants had to be native speakers of English, live in the US, correctly answer the question: "What was the name of the woman talking in the story? (Correct answer : "Sally")". Our goal was to assess the following questions: is there convergence among participants in judging Sally's utterance felicitous across stories? Does the behavior in the truth-judgment differ depending on the story (i.e. is there a main effect of the factor Story)? Does the behavior in the truth-judgment differ depending on whether it is preceded by the felicity-judgment (i.e. is there a main effect of the factor Number of test questions per participant)?

We present the results by judgment. Concerning the justification judgment, we found that participants converged in judging Sally justified in uttering 'might p' no matter whether the story was the original or the new one or whether the justification judgment was followed by the truth-judgment (the proportion of 'justified' selections was at ceiling in all conditions). Concerning the truth judgment, participants split evenly in their truth-value judgment of the same statement with the proportion of 'True' choices not statistically different from chance no matter the type of story or whether it was preceded by the 'Justification' judgment (see Fig. 1).

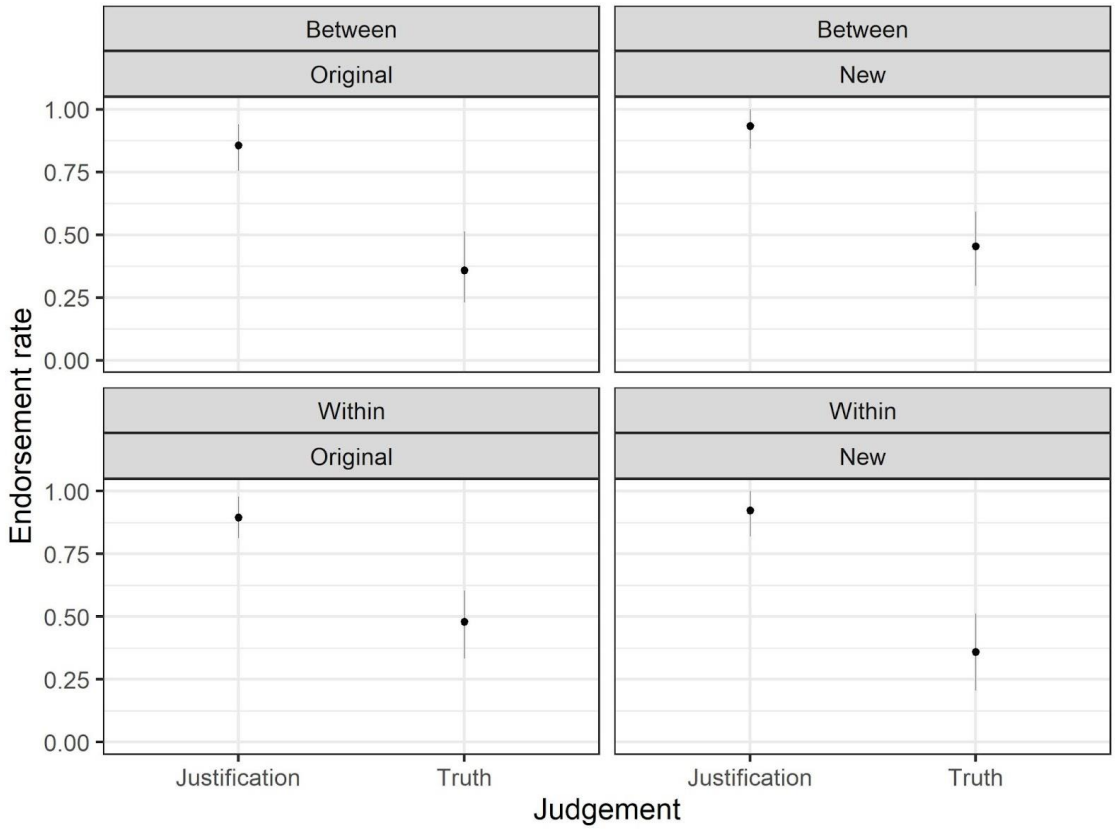

Figure 1. Experiment 1: Proportion of 'Justified' and 'True' selections by type of story and number of test questions per participants

We take our findings as suggesting that English speakers, in a situation where they all agree on judging a speaker justified to make a 'might p' statement yet split on judging whether the statement is true or false, even when they make the justification and the truth judgment one after the other. At the very least, these findings suggest that there is a significant portion of English users who judge a 'might-p' statement true as soon as they judge its speaker justified in making such a statement.

Now, notice that our 'Within' conditions in this experiment are very similar to the condition in Reuter & Brun's (2021) Experiment 3 where participants were asked in this order: "Did Maria answer to the best of her knowledge?' and 'Is Maria's answer true or false?' (See section 2 for a full description of their experiment). This allows a more direct comparison between the two experiments: in both experiments, although participants all agree that the protagonist of the story is entitled to make the statement ('might p' in our case and 'p' in Reuter & Brun's case), they disagree on whether the same statement should be judged as true or false, independently from whether the statement is 'might p' or 'p'. This raises the question as to why we didn't observe a similar behavior in the nonmodal-falsity condition in Knobe & Yalcin's (2014) Experiment 4? Experiment 2 was designed to investigate this issue.

3.2. EXPERIMENT 2. In this study, we aimed at investigating why in the nonmodal falsity condition in Knobe & Yalcin's (2014) Experiment 4 we didn't observe a disagreement among participants. Recall that the mean agreement rating for the nonmodal statement was around 6, suggesting that participants converged towards judging that what Sally said in the story (Joe is in

Boston) is false. This judgement is relevant for the discussion at stake here only if Sally is justified in making the statement based on her information. However, if one looks at the original vignette (see 3) then it seems that Sally is not justified to utter the bare 'Joe is in Boston' based on the information available to her (we are explicitly told that she doesn't know for sure), therefore there seems to be no conflict of perspective. So, it is not surprising that participants converged towards 'false' since from either the speaker's or their perspective the statement is not assertable. In this experiment we aimed at assessing two questions: (i) is our intuition correct that Sally is not justified in uttering 'Joe is in Boston' in the original vignette in Knobe & Yalcin's (2014) Experiment 4? (ii) If we modify the vignette such that Sally is justified in making the statement would we observe the same behavior observed in Reuter & Brun (2021)?

To answer question (i) we designed a task that adopted the methodology of the Within conditions of Experiment 1 relative to the bare 'p' statement in the original vignette. To answer question (ii) we manipulated the original vignette so that Sally is now justified in making the statement: in the new version Sally is uttering 'Joe is in Boston' based on a trustworthy report by Joe himself that he would have a job interview in Boston at the time of utterance but later we learn that the job interview was cancelled, and he didn't go to Boston (see 8).

(6) Experiment 2: Materials

Non-justified p (Original story)
Sally and George are talking about whether Joe is in Boston. Sally carefully considers all the information she has available and concludes that there is no way to know for sure.
Sally says: "Joe is in Boston"
Just then, George gets an email from Joe. The email says that Joe is in Berkeley. So George says: 'No, he isn't in Boston. He is in Berkeley.'

Justified p (Modified version)
It's late afternoon and Sally and George are in a pub in Berkeley talking about a common friend, Joe. Yesterday, Joe told Sally that he had a job interview in Boston at 5pm today and he would fly there early in the morning.
So, Sally says: "Joe is in Boston"
Just then, George gets an email from Joe. The email says that the job interview was cancelled and that he is still in Berkeley. So George says: "No, he isn't in Boston. He is in Berkeley."

We adopted a 2X2 design with the following two factors: SALLY'S JUSTIFICATION STATUS ('justified_p'; "nonjustified_p') and TEST QUESTION ('Justification', 'Truth'). The first factor was manipulated between-subjects while the second was manipulated within-subjects. We adopted the same procedure as Experiment 1. We recruited 100 participants (50 per story) on Prolific. We predict that in the 'justified_p' condition the proportion of 'justified' selections will be close to 1 and higher than in the 'nonjustified_p' condition, which will be close to 0. Moreover, based on the findings in Reuter & Brun (2021), we predict that the proportion of 'true' selections will be significantly higher in the 'justified_p' condition than in the 'nonjustified_p' condition.

We present the results by judgment. Concerning the justification judgment, we found that the proportion of 'justified' selection is close to ceiling in the 'justified_p' condition and close to floor in the 'nonjustified_p' condition. Concerning the truth judgment, we found that the proportion of 'true' selection is significantly higher in the 'justified_p' condition than in the

'nonjustified_p' condition as measured through a logistic regression model with "Sally's justification status" as effects-coded fixed effect (β = -1.9650, z = -2.459, p = 0.0139).

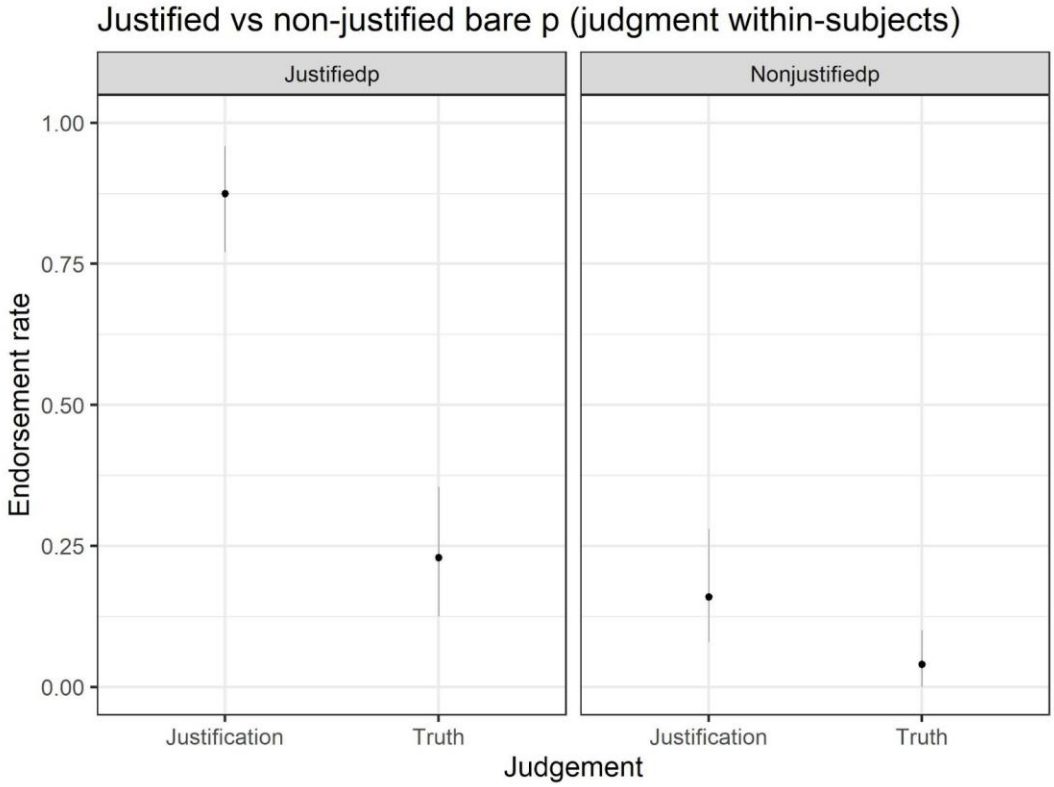

Figure 2. Experiment 2: Proportion of 'Justified' and 'True' selections by Sally's justification status.

We take these findings as suggesting that in a context where a character produces a bare 'p' statement and it is said that 'not-p', there is a higher number of participants who judge the statement 'true' when the character is justified in making the statement than when the character is not justified.

4. Discussion. In this work, we have offered findings from two experimental studies. In Experiment 1, we offered data showing that when participants converge on judging a speaker justified in making a bare 'might p' statement and they know that 'not-p', they tend to evenly split in judging whether the 'might p' statement should be labeled as 'true' or 'false'. In Experiment 2, we offered data showing that in a context where a character produces a bare 'p' statement and it is said that 'not-p', there is a higher number of participants who judge the statement 'true' when they also judge the character justified in making the statement than when they judge the character not justified. These results provide further support for an empirical generalization which was already suggested by findings reported in Knobe & Yalcin (2014) and Reuter & Brun (2021): when an agent A appropriately (based on their evidence) utters a bare 'might p' or 'p' statement but 'not-p' is the case, English speakers are not homogenous in judging whether the statement should be labeled as 'true' or 'false'. We argue that underlying this behavior is not a disagreement on the semantic content of the statements but a disagreement on how to interpret the meaning of the adjectives 'true' and 'false' when applied to a statement uttered by another agent.

Let's first consider the 'p' case. The finding is surprising assuming what can be considered the default assumption in formal semantics: as soon as one learns that 'not-p' is the case, one should judge a bare 'p' statement as false independently from the information state of the original speaker. Underlying this view are these two assumptions: (i) for a bare 'p' statement being 'true or 'false' is a matter of correspondence between the information content (the proposition) expressed by the statement and how things are in the world (in the philosophical jargon this is known as the "correspondence theory of truth"); (ii) the speaker's evidence for the proposition is not part of the proposition itself. Thus, based on these two assumptions, it is unexpected that some English speakers judge as 'true' a bare 'p' statement when in reality 'not-p' is the case. So, what is underlying this behavior? Reuter & Brun (2021) offer the following hypothesis: the lack of convergence among English speakers in performing these truth-value judgments is due to an inherent ambiguity of the adjective 'true' (and 'false') in everyday English speech. More specifically, they argue that the adjective 'true' is ambiguous between a correspondence sense (as that employed by formal semanticists) and a coherence sense according to which a "statement S of Person P [is] true or false depending on whether S coheres with other relevant beliefs of Person P, and regardless of S's correspondence with reality" (Reuter & Brun 2021: 7). Assuming this ambiguity in the use of 'true', the non-uniform behavior among participants in the relevant task can be explained as follows: those who selected 'true' adopted the coherence sense of 'true' whereas those who selected 'false' adopted its correspondence sense. Notice that using 'true' in the coherence sense is tantamount to make an assertability judgment.

Here, we suggest an alternative account of the same behavior that doesn't postulate an ambiguity for the lexical item 'true'. Recall that by default in formal semantics a statement is true if its proposition describes an actual piece of reality. At a first glance, this notion of 'actual piece of reality' seems an absolute one in the sense that establishing what counts as an actual piece of reality is independent of anyone's perspective: you just go out there in the actual world and check if the described piece of reality exists (existed or will exist) there. However, this absolute sense becomes problematic when, for example, one wants to establish the truth-value of a bare 'p' containing a predicate of personal taste like 'Jill is funny': what should one check in the actual world to assess whether the described piece of reality exists there? A common answer among semanticists is to assume that the piece of reality described by the statement 'Jill is funny' does not exist independently from some human agent's perspective and that to establish the truth of the statement it is necessary to first assume someone's perspective on the world (e.g., Lasersohn 2005, Stephenson 2007;) In other words, statements containing predicates of personal taste can never be true simpliciter but they can only be *true for someone*, namely true in the actual world as perceived by x. Now, the standard view is to assume that the notion 'true for someone' is at play only when assessing the truth of a statement describing an inherently subjective piece of reality; instead, as soon as the statement describes an inherently objective piece of reality, there is no perspective at stake because whether an objective piece of reality exists does not depend on how someone sees the world. We speculate that the data reported here show that English speakers adopt the notion of 'true for someone' even for objective statements. More specifically, we suggest that those who selected 'true' in the task interpreted the question as 'Is the underlined statement true for the speaker?' and those who selected 'false' interpreted the task as 'Is the statement true for you?', where 'for x' can be interpreted as 'based on what x knows about the world'. Thus, we take the data reported here as supporting a view of the meaning of 'true' as denoting an inherently relativized concept.

To summarize the discussion about the bare 'p' case, findings from our Experiment 2 and from Reuter & Brun (2021) showed that in situations where an agent A appropriately (based on their evidence) utters a bare objective 'p' statement but in reality 'not-p' is the case, English speakers are not homogenous in judging whether the statement should be labeled as 'true' or 'false'. Reuter & Brun (2021) proposed to account for this lack of uniform behavior by postulating an ambiguity in the lexical items 'true' and 'false' between a coherence and correspondence sense such that the split among participants can be accounted for by assuming that some people interpreted the words in their coherent sense (those who selected 'true') and some other in the correspondence sense (those who selected 'false'). We suggested an alternative account which, instead of postulating an inherent ambiguity in the meaning of the adjective 'true' (when applied to a statement), proposes that 'true' and 'false' only have the correspondence sense but are equipped with an implicit parameter defining a perspective - 'for x' - that needs to be settled before to use the adjective in a truth-value judgment task (this assumption is not ad hoc in that it is already needed for explaining the truth-evaluation behavior with a bare 'p' containing predicates of personal taste): the split among participants between selecting 'true' and 'false' can be accounted for by assuming that some people settled the parameter on the speaker's perspective (those who selected 'true') and some other on their own perspective based on the full story (those who selected 'false'). Further empirical work is needed to discriminate between the two accounts.

Let's now go back to the bare 'might p' case. The findings reported here show that in situations where an agent A appropriately (based on their evidence) utters a bare 'might objective p' statement but in reality 'not-p' is the case, English speakers behave like in the bare 'p' case: they are not uniform in judging whether the statement should be labeled as 'true' or 'false'. What does this suggest about the semantics of bare 'might p' cases? We suggest that, as in the bare 'p' case, the disagreement pertains to the truth-evaluation and not to the semantic content. Notice that a similar hypothesis has been advanced with respect to statements containing predicates of personal taste: "sentences containing predicates of personal taste are not completely objective; their truth values vary from person to person. However, this variation in truth value does not involve a variation in semantic content: if you say roller coasters are fun and I say they are not, I am negating the very content which you assert, and directly contradicting you" (Lasersohn 2005: 684). Lasersohn's idea is that the proposition communicated by uttering 'Roller coasters are fun' doesn't have a covert 'for me' component, i.e. the expressed proposition is not inherently relativized to anyone's personal taste: it is only when one wonders about the truth-value of the statement that one needs to decide first whose personal taste is relevant (or, according to Reuter & Brun, the problem of disambiguating which of the two 'true' one means). Similarly, we speculate that the proposition expressed by epistemic bare 'might p' statements like 'Joe might be in Boston now' is not inherently relativized to an agent's (or group of agents') information state: it is only when one wonders about the truth-value of the statement that one needs to decide whose information state is relevant for making the judgement. To put it more explicitly, we argue that saying 'it is possible that/might p' is not the same as saying 'based on what I know, it is possible that p/might p: only in the second case 'p' is presented as *a possibility relative to a body of information available to the speaker* whereas in the first case 'p' is presented as a *possibility relative to how things are in the world*. What does 'possibility relative to how things are in the world' mean here? To answer this question let's step back and consider the 'p' case again. When a speaker asserts a bare 'p' like 'Joe is in Boston now' what are they communicating? Intuitively, one is expressing a piece of information (a proposition) about the world that can be roughly

paraphrases as: 'the actual world is such that the event of John being home is one of its facts at the time of utterance'. Now, we assume that a reasonable and cooperative speaker would express this piece of information based on some sort of evidence supporting the occurrence of the event of Joe being in Boston now; but that's not required by the linguistic conventions: all a speaker is doing is presenting the event of Joe being in Boston as one of the facts making up the actual world without anchoring the occurrence of that event to any specific evidentiary basis. More generally, we can say that a cooperative speaker asserting a bare 'p' is describing the world as warranting the truth of the proposition encoded in 'p'; this claim is presumably based on some evidentiary support but is not presented as *depending on that evidentiary support*, unless explicitly stated so through the insertion of phrases like 'based on what I know' or 'based on my evidence'.

Let's go back to bare epistemic 'might p': when one asserts a bare epistemic 'might p' like 'Joe might be in Boston now' what is one communicating? The standard move made in the kratzerian approach is to postulate that in this case the expressed proposition includes an explicit reference to the speaker's evidentiary basis: '*based on some evidence available to me*, the actual world is such that the event of Joe being in Boston at the utterance time is a lively possibility in it' or 'there exists at least a world w' *compatible with some evidence available to me* such that the event of Joe being in Boston occurs in w' at the utterance time'. This hypothesis postulates an asymmetry between bare nonepistemic statements and bare epistemic statements such that the epistemic ones but not the nonepistemic are anchored to the speaker's evidentiary basis. We, instead, suggest that the epistemic statements behave like the nonepistemic in that the expressed proposition is a piece of information about the world without any explicit reference to the speaker's evidentiary basis: *'the actual world at the time of utterance is such that it doesn't contradict* the occurrence in it of the event of John being home at the time of utterance' or 'there is at least one world w' *compatible with how the actual world is at the time of utterance* such that the event of Joe being in Boston is a fact in w' at the utterance time'. More generally, we can say that a cooperative speaker asserting a bare 'might p' is describing the world at the time of utterance as being such that it allows the possibility of 'p' being true; presumably this claim is based on some evidentiary support (something that the speaker knows about the world now) but is not presented as depending on that evidentiary support, unless explicitly stated so through the insertion of phrases like 'based on what I know' or 'based on my evidence'. To say this in more formal terms, let's assume a standard semantic machinery where expressions are interpreted relative to a context, a world, a time, and an accessibility relation. The semantics for 'might' would be as follows:

(7) $[\![\text{might p}]\!]^{c,w,t,R} = 1$ iff $\exists w'$ such that $w' \in R_t(w)$ and $[\![p]\!]^{c,w',R} = 1$
Where $R_t(w)$ = the set of worlds identical to how w is at t

To summarize, we argue that bare 'p' and bare epistemic 'might p' are alike in expressing a proposition that describes a feature of the world without anchoring that feature to the speaker's evidentiary support.

5. Conclusion. We have showed that in situations where an agent A appropriately (based on their evidence) utters a bare 'p' or a bare 'might p' statement but in reality 'not-p' is the case, English speakers are not homogenous in judging whether the statement should be labeled as 'true' or 'false'. This behavior suggests a disagreement among English speakers on whether A's evidence at the time of utterance is relevant for the assessing the truth of a statement no matter whether a 'p' or a 'might p'. We have argued that underlying this behavior is not a disagreement

on the semantic content of the statements but a disagreement on how to interpret the meaning of the adjectives 'true' and 'false' when applied to a statement uttered by another agent. Specifically, we have discussed two potential accounts: the first account ascribes to the adjective 'true' an ambiguity between a 'coherent' sense (a statement uttered by A is true as long as the statement coheres with A's evidence) and a 'correspondentist' sense (a statement as uttered by A is true as long as the statement correspond to reality); the second account treats 'true' as having only the correspondentist sense but ascribes to it an inherent relative nature (to decide if a statement as uttered by A is true one has to first decide a perspective, i.e. A's or the assessor's). Further research is needed to discriminate between these two accounts. Lastly, we have discussed the implications of the findings for the semantics of epistemic 'might p' statements: in our interpretation these findings suggest that the semantic content of bare 'might p' statements is not inherently relativized to an agent's information state but the relativization takes place only when one assesses the truth-value of the statement. This hypothesis treats bare 'p' and bare 'might p' symmetrically: in both cases the expressed proposition is not implicitly relativized.

References

Doran, Ryan, Gregory Ward, Meredith Larson, Yaron McNabb & Rachel E. Baker. 2012. A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language* 88(1). 124–154. https://doi.org/10.1353/lan.2012.0008.

Dowell, J.L. 2011. A flexible contextualist account of epistemic modals. *Philosophers' Imprint* 11(14). 1–21. http://hdl.handle.net/2027/spo.3521354.0011.014.

Egan, Andy. 2007. Epistemic modals, relativism, and assertion. *Philosophical Studies* 113(1). 1–22. https://doi.org/10.1007/s11098-006-9003-x.

Egan, Andy, John Hawthorne & Brian Weatherson. 2005. Epistemic modals in context. In G. Preyer & G. Peter (eds.), *Contextualism in philosophy*, 131–170. Oxford: Oxford University Press.

von Fintel, Kai & Anthony Gillies. 2008. CIA leaks. *Philosophical Review* 117(1). 77–98.

von Fintel, Kai & Anthony Gillies. 2011. 'Might' made right. In Andy Egan & BrianWeatherson (eds.), *Epistemic modality*, 108–130. Oxford: Oxford University Press.

Jasbi, Masoud, Brandon Waldon, & Judith Degen. 2019. Linking hypothesis and number of response options modulate inferred scalar implicature rate. *Frontiers in Psychology* 10. 189. https://doi.org/10.3389/fpsyg.2019.00189.

Khoo, Justin & Jonathan Phillips. 2019. New horizons for a theory of epistemic modals. *Australasian Journal of Philosophy* 97(2). 309–324.

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. 1977. What 'must'and 'can'must and can mean. *Linguistics and Philosophy* 1(3). 337–355. https://doi.org/10.1007/BF00353453.

Kratzer, Angelika. 1981. The notional category of modality. In Hans-Jürgen Eikmeyer & Hannes Riese (eds.) *Words, worlds, and contexts*, 38-74. Berlin: de Gruyter.

Kratzer, Angelika. 1991. Modality. In Armin Von Stechow & Dieter Wunderlich (eds.) *Semantics: An international handbook of contemporary research*, 639–650. Berlin: de Gruyter

Kratzer, Angelika. 2012. *Modals and conditionals: New and revised perspectives*. Oxford: Oxford University Press.

Lasersohn, Peter. 2005. Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy* 28(6). 643–686. https://doi.org/10.1007/s10988-005-0596-x.

MacFarlane, John. 2011. Epistemic modals are assessment-sensitive. In Andy Egan & Brian Weatherson (eds.), Epistemic modality, 144–178. Oxford: Oxford University Press.
MacFarlane, John. 2014. *Assessment sensitivity*. Oxford: Oxford University Press
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.
Phillips, Jonathan & Matthew Mandelkern. 2020. Eavesdropping: What is it good for? *Semantics and Pragmatics* 13(19). 1–30. https://doi.org/10.3765/sp.13.19.
Reuter, Kevin & Georg Brun. 2021. Empirical studies on truth and the project of re-engineering truth. *Pacific and Philosophical Quarterly.* 1-25
Rudin, Deniz. 2021. Asserting epistemic modals. Manuscript, University of Southern California.
Scontras, Gregory & Lisa S. Pearl. 2021. When pragmatics matters more for truth-value judgments: An investigation of quantifier scope ambiguity. *Glossa: A Journal of General Linguistics* 6(1). 1–37
Stalnaker, Robert. 2014. *Context*. Oxford: Oxford University Press
Stephenson, Tamina. 2007. Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy* 30. 487–525. https://doi.org/10.1007/s10988-008-9023-4.
Tonhauser, Judith & Lisa Matthewson. 2015. Empirical evidence in research on meaning. Manuscript.
Waldon, Brandon & Judith Degen. 2020. Modeling behavior in truth value judgment task experiments. *Proceedings of the Society for Computation in Linguistics* 3(1). 10–19.
Yalcin, Seth. 2007. Epistemic modals. *Mind* 116. 983–1026.
Yalcin, Seth. 2011. Nonfactualism about epistemic modals. In Andy Egan & Brian Weatherson (eds*.), Epistemic modality*, 295–332. Oxford: Oxford University Press.
Yanovich, Igor. 2014. Standard contextualism strikes back. *Journal of Semantics* 31. 67–114. https://doi.org/10.1093/jos/ffs022.