Materials for Eavesdropping: What is it good for?

Jonathan Phillips & Matthew Mandelkern

May 2020

Knobe & Yalcin 2014: Falsity and Retraction

Participants

We collected a sample of 242 participants ($M_{age} = 37.09$; $SD_{age} = 11.3$; 104 females) from Amazon Mechanical Turk (www.mturk.com).

Procedure

Participants were randomly assigned to one of four conditions, two of which were exact replications of the conditions in Experiment 5 in (Knobe & Yalcin, 2014). As in (Knobe & Yalcin, 2014), all participants read a version of the following scenario, which varied only in what exactly Sally said. For example, in the non-modal variant, Sally makes a non-modal claim about Joe's location.

> *Non-modal variant*: 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."

In the *Epistemic modal variant*, Sally instead says "Joe might be in Boston", and critically in the *Attitude report variant*, Sally says "I think Joe is in Boston." No other changes were made to the original materials.

After reading one of these variants, participants either asnwered a question about whether it would be appropriate for Sally to retract what she said or answered a question about whether what Sally said was false:

> *Retraction question*: We want to know whether it would be appropriate for Sally to take back what she said (for example, by saying "Ok, scratch that"). So please tell us whether you agree or disagree with the following statement:
>
> - It would be appropriate for Sally to take back what she said.
>
> *Falsity question*: 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 was false.

In both cases, participants indicated their agreement on a scale from 1 ('Completely disagree') to 7 ('Completely agree'). Finally, all participants completed a brief demographic questionnaire.

Results

Replication

To statistically characterize the pattern of responses (see Fig. 1), we first asked whether we replicated the original finding in (Knobe & Yalcin, 2014).[1] Similar to (Knobe & Yalcin, 2014), we observed the critical

[1] An overall analysis of variance revealed that participants' agreement ratings were significantly affected by whether the agent uttered a non-modal assertion, an epistemic modal claim, or an attitude report, $F = 68.2$, $p < 0.001$, $\eta_p^2 = 0.17$. We also

interaction effect between question-type (Retraction vs. Falsity) and statement-type (Non-modal vs. Modal), F (1,156) = 34.64, $p < 0.001$, $\eta_p^2 = 0.182$. This interaction was driven by the fact that we observed no significant difference between judgments of falsity and retraction for non-modal claims, $t(66.38)$ = -1.62, $p = 0.109$, $d = 0.363$, but did observe a large difference between falsity and retraction judgments for epistemic modal claims, $t(64.99)$ = -7.64, $p < 0.001$, $d = 1.689$.

Extension

We next asked whether the observed pattern for epistemic modal claims extended to attitude reports. Indeed, we found a similar pattern: when the agent's claim involved an attitude report rather than an epistemic modal claim, participants were more likely to judge that the agent should retract the claim than that the claim was false $t(66.91)$ = -4, $p < 0.001$, $d = 0.883$. Moreover, we found no significant difference in participants' agreement that retraction would be appropriate when the the claim involved an epistemic modal or an attitude report, $t(79) = 1.06$, $p = 0.291$, $d = 0.236$, and found that, if anything, participants more agreed with the falsity of the claim when it involved an attitude report than when it involved an epistemic modal, $t(80)$ = -2.449 $p = 0.016$, $d = 0.541$.

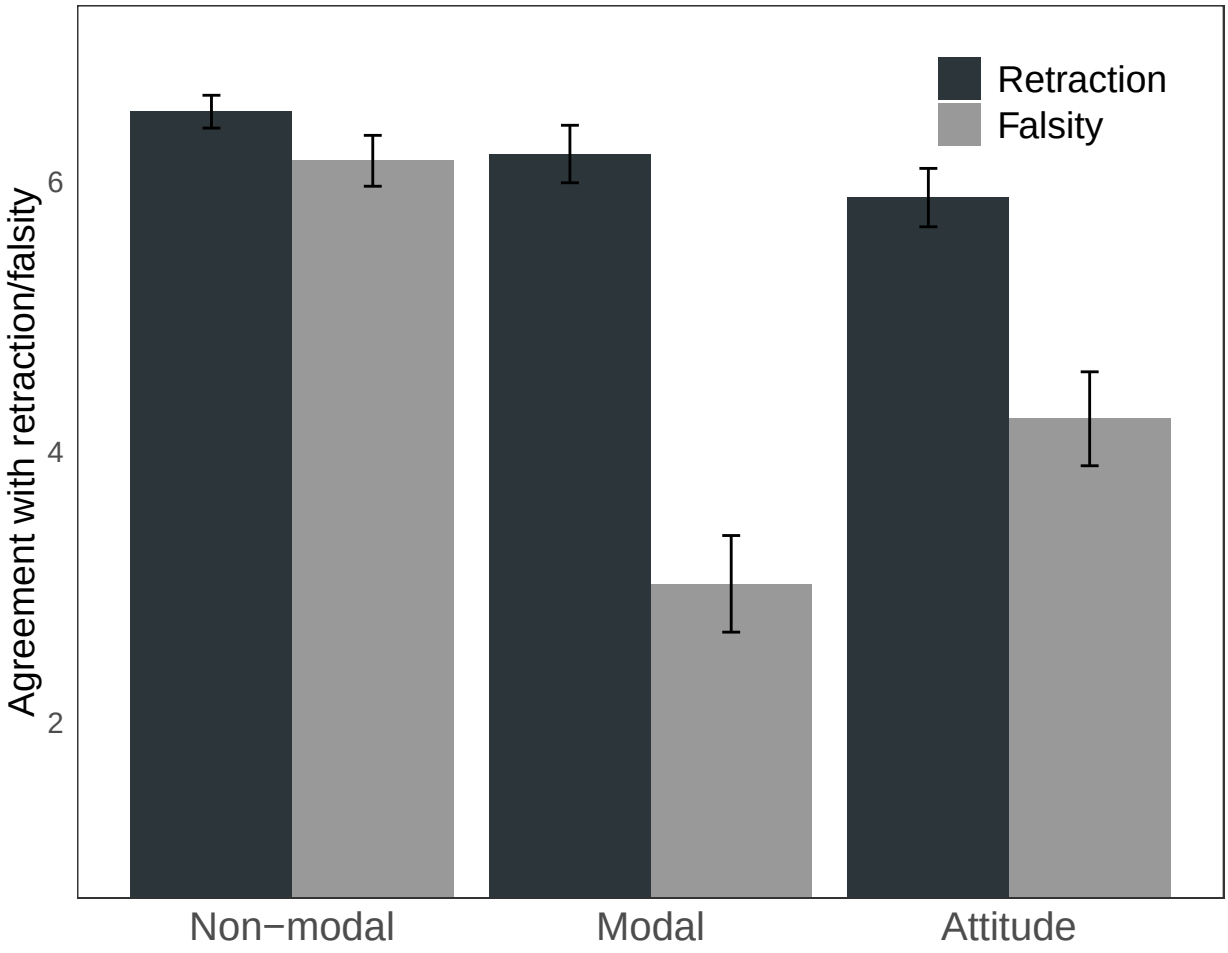

Figure 1: Participants' mean level of agreement with the retraction or falsity claims. Errors bars indicate +/- 1 *SEM*.

observed a significant effect of whether participants were asked about whether it would be appropriate for the agent to retract her claim or whether the agent's claim was false, $F = 24.12$, $p < 0.001$, $\eta_p^2 = 0.224$. More importantly, we also observed an interaction between these two variables, $F = 14.87$, $p < 0.001$, $\eta_p^2 = 0.112$, meaning that the pattern of participants' judgments about the different claims differed depending on whether they were asked the retraction or falsity question.

Beddor & Egan 2018: Falsity and QUDs

Participants

We collected a sample of 500 participants (M_{age} = 38.33; SD_{age} = 13.21; 240 females) from Amazon Mechanical Turk (www.mturk.com).

Procedure

Participants were randomly assigned to one of four conditions, two of which were exact replications of the conditions in Experiment 5 in (Beddor & Egan, 2018). As in (Beddor & Egan, 2018), all participants in these conditions began by reading the following preamble:

> *Preamble*: John is worried he might have strep throat. He goes to his primary care physician and she runs an initial test that indicates that there is a 75% chance that John does not have strep. Based on the initial test results, John's doctor says: "You probably don't have strep throat."

Then participants were randomly assigned to one of two conditions which continued this. In one condition, the question under discussion was meant to target the prejacent of the doctor's utterance:

> *QUD-Prejacent*: John comes back two days later to find out the results of the throat culture, and sees a different doctor. The throat culture comes up positive, which indicates there is a 90% chance that John has strep throat. John has not yet seen the results of these tests, but his new doctor has.
>
> John asks the new doctor: "I'm trying to figure out whether I need to take antibiotics. My primary care physician told me, 'You probably don't have strep.' Is what she said true?"

Participants then answered the question, "Which of the following responses would be correct?" by selecting either "No, it's not" or "Yes, it is".

In the other condition, the question under discussion targeted the doctor's competence:

> *QUD-Competence*: John comes back two days later to find out the results of the throat culture, and sees a different doctor. The throat culture comes up positive, which indicates there is a 90% chance that John has strep throat.
>
> But now John wants to know whether his primary care physician made a mistake administering the initial test, so he asks: "I'm trying to figure out whether I can rely on my primary care physician. She told me, 'You probably don't have strep'. Is what she said true?"
>
> The new doctor reviews the initial tests, and confirms that John's primary care physician had not made any mistakes interpreting the results.

Participants then answered the question, "Given this, which of the following responses would be correct?" by selecting either "No, it's not" or "Yes, it is".

As in (Beddor & Egan, 2018), participants then completed a comprehension check question: "In the scenario you just read, which of the following did John's primary care physician say?" The options were (a) You probably have strep. (b) You probably don't have strep. (c) You probably have pneumonia. (d) You probably don't have pneumonia.

In the other two conditions, which extended those of (Beddor & Egan, 2018), we simply replaced the doctor's utterance, "You probably don't have strep throat" with the utterance "I don't think you have strep throat". Accordingly, we also changed the comprehension check question such that the options were: (a) I think you have strep. (b) I don't think you have strep. (c) I think you have pneumonia. (d) I don't think you have pneumonia. No other changes were made to the original materials.

Finally, all participants were asked to write a complete sentence about what the study was about and then completed a brief demographic questionnaire.

Results

We excluded the 27 participants who failed to answer the comprehension check question correctly, and analyzed the responses from the remaining 473 participants. Replicating (Beddor & Egan, 2018), when the utterance involved the term 'probably', we found that 83.05% of participants in the *QUD-Prejacent* condition selected "No, it's not", while only 28.81% of the participants in the *QUD-Competence* condition did.

More importantly, we found a strikingly similar pattern of respons when the utterance was instead about the what the doctor believed. We again found that 74.79% of participants in the *QUD-Prejacent* condition selected "No, it's not", while only 27.12% of the participants in the *QUD-Competence* condition did (Figure 2).

Moreover, analyzing all of the data using a genearlized linear model, we found only a main effect of the QUD manipulation, $z = 7.826$, $p < 0.001$, and no effect of whether the utterance involved an epistemic modal or an interaction between them, $p \geq 0.121$.

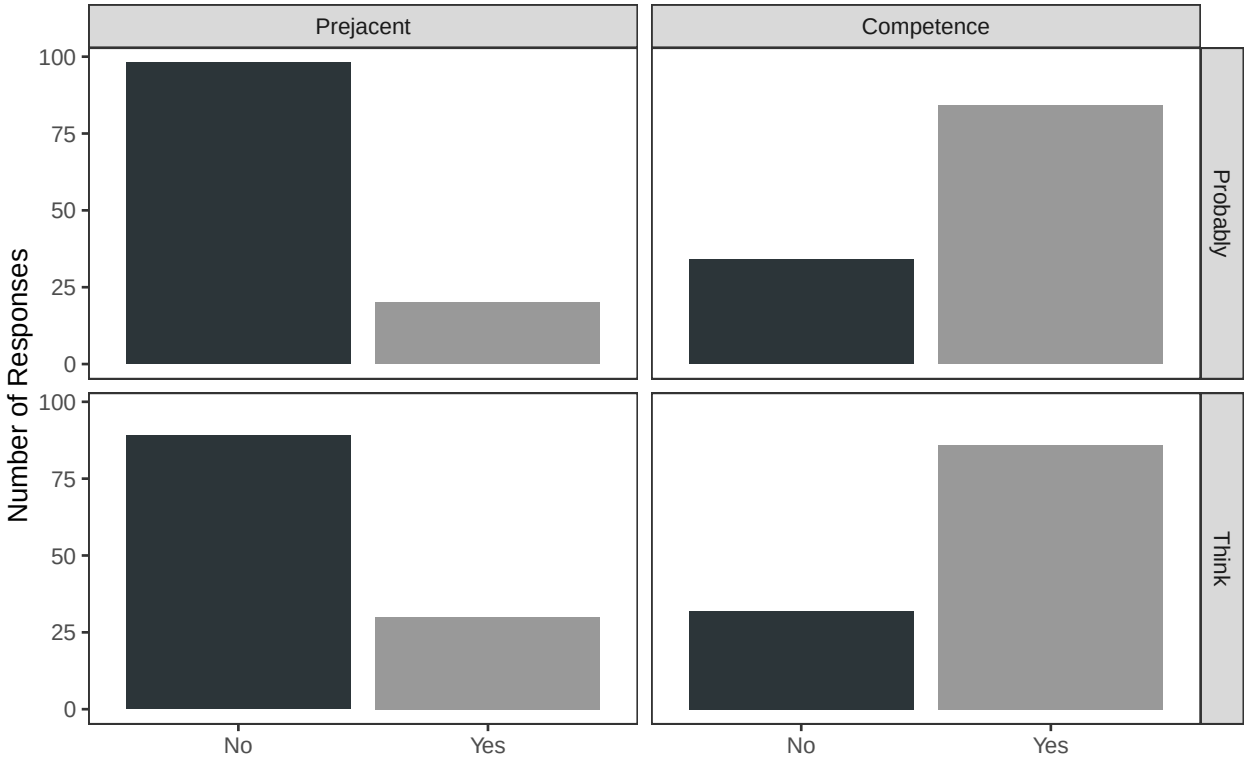

Figure 2: Participants' responses as a function of both whether the utterenace involved an espistemic modal term and whether the QUD intended to target the prejacent or the doctor's competence.

Khoo & Phillips 2019: Consistency

Participants

In this experiment, 405 participants (M_{age}=37.99, SD_{age}=12.47; 192 females) were recruited through Amazon Mechanical Turk (http://www.mturk.com).

Procedure and Materials

Following (Khoo & Phillips, 2019) Participants completed a single trials which each involved reading a vignette about an ongoing police investigation. In all cases, participants first read the following **background** information:

> 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.

How this background continued dependend on the condition to which participants wre randomly assigned. Participants were assigned to either an **Utterances** or an **Assessments** condition, and additionally to make assessments of either epistemic **Modal** claim, a **Non-Modal** claim, or and **Indexical** claim.

In the **Modal Utterances** case, the background continued as follows:

> 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 case**, the background instead continued:

> 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."**

Two other conditions differed slightly from these. In one, 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 instead read: "Fat Tony died [did not die] at the docks." In the other, participants instead evaluated an **Attitude** claim. These cases were identical to the preceding ones except that the Inspectors'/Chief's claim(s) took the form of an attitude report, and thus instead read: "I [think / don't think] Fat Tony died at the docks."

Finally, other pariticpants instead assessed **Indexical** statements. In the **Indexical Utterances** case, the **Background** continued as follows:

> 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."

In the **Indexical Assessments** condition, the two inspectors instead made two different claims about the truth of the Chief's utterance:

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."

Questions

After reading the entire vignette, 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"").

After answering this question, participants also answered a manipulation check question. In the **Modal**, **Non-Modal**, and **Attitude** conditions, participants 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 test 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 the **Indexical** conditions, participants were instead separately asked who both Inspector A and Inspector B think has served on the police force for twenty years. They responded by selecting one of the following three options for each Inspector:

- Inspector A
- Inspector B
- The Chief

Finally, participants completed a brief and optional demographic questionnaire which included questions about their age, gender, ethnicity, education, and SES.

Manipulation Check Questions

We first analyzed 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, $\chi^2(1) = 153.4$, \$p < \$ 0.001, $V = 0.5$.

Replication

Following (Khoo & Phillips, 2019), 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 ($M = 3.42$, $SD = 1.94$). 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 ($M = 5.65$, $SD = 1.68$), $t(94) = -6.02$, $d = 1.23$.

Next, we analyzed participants' compatibility judgments in the **Modal** and **Non-Modal** conditions with a 2 (Statement: Bare vs. Modal) × 2 (Condition: Utterances vs. Assessments) ANOVA. Replicating (Khoo & Phillips, 2019), participants' judgments were significantly affected by whether or not the claims involved a bare assertion or an epistemic modal claim, F (1,199) = 17.95, $p < 0.001$, $\eta_p^2 = 0.083$. More specifically, we found that participants more strongly agreed that one of the inspectors' claims must be false when they uttered/assessed a bare assertion ($M = 5.69$, $SD = 1.46$), than when they uttered/assessed a modal claim ($M = 4.62$, $SD = 2.06$), $t(184.22) = -4.27$, $p = 0.001$, $d = 0.6$. As in (Khoo & Phillips, 2019), we also did not observe a significant effect of whether the Inspectors made conflicting utterances or conflicting assessments, $F(1,199) = 0.29$, $p = 0.588$, $\eta_p^2 = 0.001$, and did not find an interaction effect between these two variables, $F(1,199) = 0.09$, $p = 0.764$, $\eta_p^2 < 0.001$, meaning that the difference between the different claims (**Bare** vs. **Modal**) did not significantly differ between the **Assessments** and **Utterances** conditions.

Extension

We then asked whether the pattern we observed in the modal condition could similarly be found in the attitude report condition. Specifically, we did a similar analysis to that in (Khoo & Phillips, 2019) but replaced the modal condition with the attitude condition. Once again, we found that participants' judgments were significantly affected by whether or not the claims involved a bare assertion or an attitude report, F (1,202) = 4.77, $p = 0.03$, $\eta_p^2 = 0.023$. Specifically, we found that participants more strongly agreed that one of the inspectors' claims must be false when they uttered/assessed a bare assertion ($M = 5.69$, $SD = 1.46$), than when they uttered/assessed an attitude report ($M = 5.16$, $SD = 1.95$), $t(194.12) = -2.21$, $p = 0.028$, $d = 0.31$. We again did not observe a significant effect of whether the Inspectors made conflicting utterances or conflicting assessments, $F(1,202) = 0.05$, $p = 0.82$, $\eta_p^2 < 0.001$, and did not find an interaction effect between these two variables, $F(1,202) = 0.21$, $p = 0.645$, $\eta_p^2 = 0.001$, meaning that the difference between the different claims (**Bare** vs. **Attitude**) did not significantly differ between the **Assessments** and **Utterances** conditions.

Finally, the **Modal** and **Attitude** conditions did not differ significantly from one another, $t(205.63) = -1.94$, $d = 0.27$.

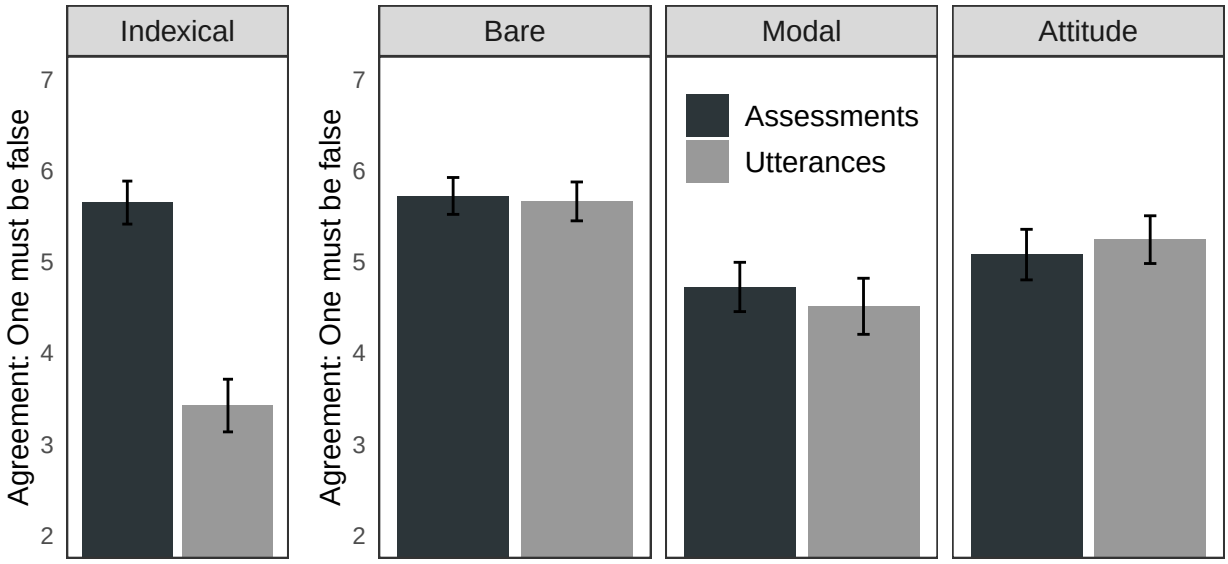

Figure 3: Participants' mean level of agreement that at least one of the inspectors' claims must be false. Errors bars indicate +/- 1 *SEM*.

Beddor, B., & Egan, A. (2018). Might do better: Flexible relativism and the qud. *Semantics and Pragmatics*, *11*.

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

Knobe, J., & Yalcin, S. (2014). Epistemic modals and context: Experimental data. *Semantics and Pragmatics*, *7*, 10–11.