

The Weakness of Epistemic *must*

Gregory Scontras (scontras@stanford.edu), Judith Degen (jdegen@stanford.edu),
Justine Kao (justinek@stanford.edu), Erin Bennett (erindb@stanford.edu),
Noah D. Goodman (ngoodman@stanford.edu)

Department of Psychology, Stanford University
Stanford, CA 94305 USA

Abstract

We present results from three studies exploring the interpretation of statements featuring the epistemic necessity modal *must*, comparing these statements to utterances with weaker modals, or no modals at all. Our results demonstrate that *must*-statements coincide with less certain belief states, such that “It must be raining” receives a weaker interpretation than bare “It is raining.” Rather than engineering weakness into the meaning of the word *must*, our account derives its weakness as an M-implicature (Levinson, 2000). Statements with *must* are marked relative to bare statements, so they communicate a marked meaning, namely less certainty on the part of the speaker. A general model of rational inference in communication delivers the puzzlingly weak interpretation of *must*.

Keywords: pragmatics; semantics; psycholinguistics; modals; M-implicature

Introduction

Modals come in varying degrees of strength; the strongest are modals of necessity like *must* or *have to*. Under a quantificational treatment of modality, these necessity modals correspond to universal quantifiers over possible worlds (XXX citation). When used, they assert that in every (relevant) possible world, some proposition *p* holds. In (1b), epistemic *must* is used to assert that in every world compatible with the speaker’s knowledge, it is raining. Given that knowledge corresponds to justified true belief (that which is known cannot be otherwise), from (1b) it necessarily follows that it is raining (i.e., it is not possible that it is not raining). There are other analyses of modals on the market (e.g., a scalar semantics for modals; Lassiter, 2011), but all of them share the intuitive understanding of necessity modals as maximally strong semantic objects.

- (1) a. It might be raining.
- b. It must be raining.
- c. It is raining.

When compared with other modals, necessity modals do behave as strong: (1b) expresses a stronger statement about rain than (1a). However, when compared to the bare statement without a modal, *must* and other epistemic necessity modals suddenly behave as weak, or so the liturgy goes. At issue is the failed inference from (1b) to (1c) first observed by Karttunen (1972): How could *must p* not entail *p*? If it is necessarily raining, then surely it is raining. It appears that talking about what is necessarily the case commits speakers to less than does talking about what is actually the case. Karttunen and decades of semanticists that followed posit that the inference from (1b) to (1c) fails because *must p* is

a weaker statement than bare *p*. In other words, *must* is weak. Our first task is to evaluate the relative strength of *must*, an empirical investigation conspicuously absent from the discussion of this modal thus far. Having justified the strength of *must* empirically, we may then set our sights on the proposals meant to account for *must*’s meaning.

Linguists split on the semantics of *must*. The original “*must* is weak” mantra, stemming from Karttunen (1972), has *must p* not entail *p*. There are two tacks to breaking this entailment. For Kratzer (1991), the entailment fails because *must p* is weaker than we might have thought. She weakens *must* by having it quantify not just over knowledge states, but also over expectations given knowledge states – expectations which are not always borne out. For Veltman (1985), the entailment fails because bare *p* is stronger than we might have thought. He strengthens bare *p* by allowing partial knowledge states, such that *must p* may be true while we lack the knowledge to say whether *p* is true or false. Either way, from *must p* (e.g., “It must be raining”) it does not follow that *p* (e.g., “It is raining”).

von Fintel and Gillies (2010) lead the opposing charge, holding firm to the logic that *must p* is strong, quantifying universally over epistemically possible worlds. In addition to this strong semantics, they also claim that *must* is an evidential marker, presupposing that the speaker has no direct evidence of *p*. However, the strength of *must p* requires the speaker to have direct evidence of something that entails *p*. Using a corpus of naturally-occurring examples from Ancestry.com, Lassiter (*to appear*) shows how von Fintel and Gillies’s implementation of a strong semantics for *must* makes unreasonable claims about the knowledge states of speakers. Lassiter proposes instead a weak probabilistic semantics: *must p* entails that the probability of *p* given the speaker’s direct knowledge is greater than chance, and requires that the question of whether *p* not be resolved by this direct knowledge. Thus, he returns to the original mantra of weakness while building evidential information into the semantics of *must*.

In an effort to inform this debate on the strength of *must*, we first investigate what gets communicated via its use. Over the course of three experiments, we assess the speaker commitments, evidence strength, and listener understanding that accompany *must* statements. Our findings qualitatively confirm Karttunen’s original observation, and quantitatively demonstrate the relative weakness of *must*. Moreover, they highlight the role of evidence strength in the computation of *must*’s meaning – an ingredient that has thus far been lacking from

the proposed semantics for *must*.

We then use our findings to motivate a computational model of the meaning of *must*, following Lassiter and Goodman (2013) in their extension of the Bayesian Rational Speech Act framework (Frank & Goodman, 2012). Rather than engineering weakness relative to bare *p* directly into the meaning of the word *must*, our account derives its weakness as an M-implicature (Levinson, 2000): *must p* is marked (i.e., costly) relative to the bare form (1c); the bare form is sufficiently strong already to convey *p* (e.g., that it is raining), so listeners take the marked form to convey the marked meaning that the speaker arrived at the conclusion *p* via an evidentially less certain route than if they had chosen the shorter, bare form. Our account accomplishes two things. First, it obviates the conceptual pitfalls of von Stechow and Gillies’s strong semantics, without treating *must* as inherently weak: evidence strength is inferred from, rather than communicated by, the use of *must*. Second, it implements a formal model of M-implicatures within the Rational Speech Act framework (Bergen, Levy, & Goodman, 2014).

Experiment 1: Evidence

Most discussions of *must* make claims about the quality and/or quantity of evidence involved in its assertion. To better understand the meaning of *must*, we therefore began by collecting estimates of evidence strength. These estimates were used in the analyses of Experiments 2 and 3, which target *must*’s meaning directly.

Participants, materials, and methods

40 participants were recruited through Amazon’s Mechanical Turk crowd-sourcing service, and were compensated for their participation.

Subjects rated the probability of some state of affairs *p* (e.g., “It is raining”) given a piece of evidence (e.g., “You hear the sound of water dripping on the roof”) on a sliding scale with endpoints labeled “impossible” and “absolutely certain”. There were four difference states (rain, cold coffee, dinner ready, dog barking), and for each state subjects evaluated one of five possible pieces of evidence.

Results and discussion

Probability ratings were averaged across subjects for each state of affairs and evidence type, and ranged from XX to XX, with a standard deviation of XX. These average ratings serve as a measure of evidence strength: higher probability ratings signal greater strength of evidence. We used these ratings in the design and analyses of Experiments 2 and 3.

Experiment 2: Speakers

Next, we evaluated the intuitions of speakers in a forced production task, testing how likely they are to use *must p* vs. bare *p* as evidence strength decreases. In other words, we tested the relationship between *must* and the evidence justifying its use.

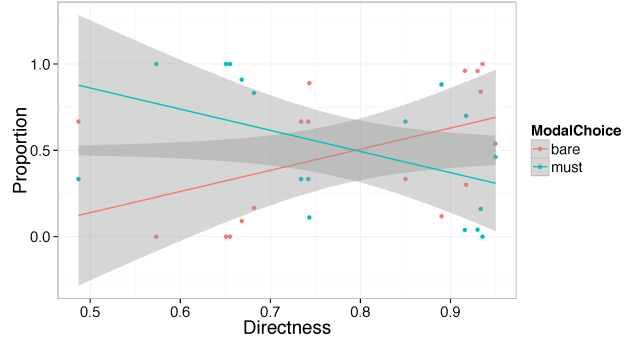


Figure 1: Utterance choice plotted against evidence strength (Expt. 2).

Participants, materials, and methods

We recruited 40 participants from Amazon’s Mechanical Turk. Participants were compensated with a small payment.

On each trial, participants were presented with a piece of evidence (e.g., “You see a person come in from outside with wet hair and wet clothes”) and were asked to choose one of three possible utterances to describe a state of affairs *p* to a friend given the available evidence. We used the same pieces of evidence and states of affairs from Experiment 1. For example, in describing the rain state, speakers chose between the three utterances in 2.

- (2) a. It might be raining. (*might*)
- b. It must be raining. (*must*)
- c. It is raining. (*bare*)

Speakers completed four trials, one for each of the possible states.

Results and discussion

Fig. 1 plots selection rates for the *might*, *must*, and *bare* utterances against the evidence strength estimates from Experiment 1. Participants were more likely to choose the more marked *must* form over the bare form as evidence strength decreased ($\beta=5.4$, $SE=2.4$, $p<0.05$).

Experiment 3: Listeners

Having found that speakers’ use of *must* increases as evidence strength decreases, we next tested whether listeners take into account this information about speakers as they interpret the bare and *must* forms. In testing listeners’ interpretation of *must*, we also evaluated its relative strength.

Participants, materials, methods

We recruited 120 participants through Amazon’s Mechanical Turk. Participants were compensated with a small payment.

We used the same four states of affairs and pieces of evidence from Experiments 1 and 2. On each trial, participants were presented with an utterance (e.g., “It must be raining”) and asked a) to rate the probability of the state of affairs *p*

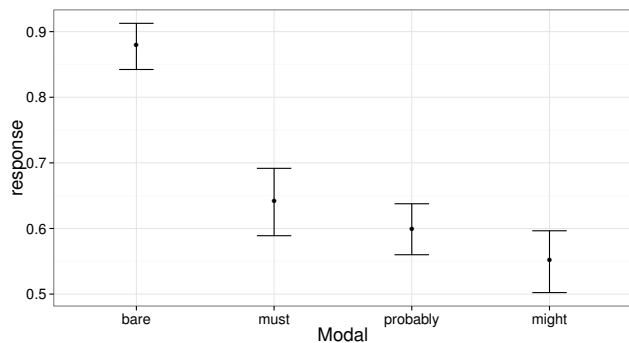


Figure 2: Listener beliefs about the state of affairs p after hearing a given utterance (Expt. 3).

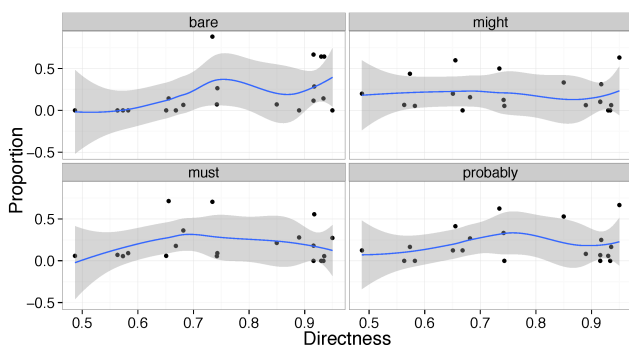


Figure 3: Assumed evidence strength for a given utterance (Expt. 3).

(e.g., of rain) on a sliding scale with endpoints labeled “impossible” and “certain”; and b) to select one out of five pieces of evidence that the speaker had about p in making their utterance.

Results and discussion

Fig. 2 plots listener belief given utterance: participants believed p was less likely after observing the *must* utterance than after observing the *bare* utterance ($\beta=-0.21$, $SE=0.01$, $t=-10.1$, $p<0.0001$). Fig. 3 plots average strength of evidence assumed for each utterance. Average evidence strength was lower for *must* than for the bare utterance ($\beta=-0.08$, $SE=0.01$, $t=-6.8$, $p<0.0001$).

Model

We propose a computational model of language understanding to show that the empirically verified, relatively weak interpretation of *must* does not require encoding weakness or indirectness into the semantics of modals, but can arise instead from pragmatic reasoning about its strength. In fact, already in Grice (1989, pp. 33–34) do we find inspiration for the calculation that leads to the relative weakness of *must*:

“A wants to know whether p , and B volunteers not only the information that p , but information to the effect

that it is certain that p ... B’s volubility may be undesigned, and if it is so regarded by A it may raise in A’s mind a doubt as to whether B is as certain as he says he is... But if it is thought of as designed, it would be an oblique way of conveying that it is to some degree controversial whether or not p .”

Here, we formalize Grice’s description of the computation.

Our model follows the basic structure of Rational Speech Act (RSA) models, which view language understanding as recursive reasoning between speaker and listener (Frank & Goodman, 2012).

Conclusion

Acknowledgments

References

- Bergen, L., Levy, R., & Goodman, N. D. (2014). *Pragmatic reasoning through semantic inference*. (Unpublished MS, Available online at <http://web.mit.edu/bergen/www/papers/BergenLevyGoodman2014.pdf>)
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336, 998–998.
- Grice, P. (1989). *Studies in the way or words*. Cambridge, MA: Harvard University Press.
- Karttunen, L. (1972). *Possible and must*. In J. Kimball (Ed.), *Syntax and semantics* (Vol. 1, p. 1–20). New York: Academic Press.
- Kratzer, A. (1991). Modality. In A. von Stechow & D. Wunderlich (Eds.), *Semantics: An international handbook of contemporary research* (p. 639–650). Berlin: de Gruyter.
- Lassiter, D. (2011). *Measurement and modality: The scalar basis of modal semantics*. Unpublished doctoral dissertation, New York University.
- Lassiter, D. (to appear). The weakness of *must*: In defense of a mantra. In T. Snider (Ed.), *Proceedings of SALT 24*.
- Lassiter, D., & Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. In T. Snider (Ed.), *Proceedings of SALT 23* (p. 587–610). CLC Publications.
- Levinson, S. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Veltman, F. (1985). *Logics for conditionals*. Unpublished doctoral dissertation, University of Amsterdam.
- von Fintel, K., & Gillies, A. S. (2010). *Must ... stay ... strong!* *Natural Language Semantics*, 18, 351–383.