Epistemic Meaning and the LLL Tune in American English

Thomas Sostarics[1], *Jennifer Cole*[1]

[1]Department of Linguistics, Northwestern University

tsostarics@u.northwestern.edu, jcole.1@northwestern.edu

Abstract

In American English, the pitch pattern at the end of a prosodic phrase—the nuclear tune—conveys pragmatic meaning, which can be characterized in epistemic terms as relating to the mutual beliefs of the speaker and their addressee [1]. Epistemic accounts have been offered for the meaning of rising and falling nuclear tunes [2], or a mid-level pitch plateau [3], but the epistemic contribution for many other tunes has not been addressed. This paper addresses one such gap in the literature, examining the meaning of the L*L-L% contour in American English through two experiments testing listeners' preferences for H*L-L% vs. L*L-L% with declarative sentences as a function of whether the associated proposition is ruled in (congruent with speaker and addressee shared knowledge) or ruled out (incongruent). Our results show that listeners find L*L-L% to be more felicitous than H*L-L% in ruled-out contexts, despite prior claims that H*L-L% is the default tune for declarative assertions. A follow-up experiment shows that the preference for L*L-L% in ruled-out contexts remains even when the intonation is made redundant by explicit acknowledgement of the ruled-out status of the proposition. We conclude that L*L-L% makes an independent epistemic contribution to utterance meaning in American English and propose that a speaker uses L*L-L% to signal that they withhold their commitment to the propositional content of their utterance.

Index Terms: epistemic meaning, speech perception, intonation, intonational meaning

1. Introduction

Seminal work on intonational meaning in American English [1] proposes that the meaning expressed by the sentence- or phrase-final intonation pattern—the nuclear tunes—is dependent on the discourse context and mutual beliefs between a speaker and their interlocutors. Some tunes have received more attention than others in work using anecdotal examples [1, 4, 5], but rigorous empirical tests of tune meaning are limited in number [3, 6]. This gap is addressed with an empirical investigation of tune meaning for the sustained low tune, L*L-L% in the Autosegmental-Metrical model [7], phonetically implemented with a low and flat or low and falling pitch sustained over the final portion of the sentence. The L*L-L% tune has received little attention in the literature, with [8] anecdotally describing it as scathing intonation used with echo statements.

A frequently discussed tune is H*L-L%, with a falling pitch. Under a compositional account of tune interpretation, the meaning of H*L-L% is comprised of the meanings of its constituent tones where the H* pitch accent denotes new information to be predicated; the L- phrase accent emphasizes the separation of the current phrase with other phrases [1]; and the L% boundary tone denotes completion or lack of a larger grouping [5]. Combined, the meaning of the H*L-L% tune can be described as adding new information to the common ground

that needs no additional followup to be interpreted. Under this account, H*L-L% is well-suited for its default association with assertive force [9, p. 90]. It is more difficult to provide a similar account for the meaning for L*L-L%. In contrast to the H* pitch accent in H*L-L%, the L* pitch accent is described as conveying the discourse-given status of a referent or the belief that a proposition p is false [10]. These proposed meanings seem at odds with the assertive function of declarative sentences, raising the question of why a speaker would ever choose L*, rather than H*, as the nuclear accent with the L-L% edge tone sequence.

Consider the discourse context in (1) where the default tune for A's response is H*L-L%. Under the proposal in [1], H*L-L% fulfills the predication x of B's question $\mathrm{GET}(A, x)$ with new information: x=oranges. If L*L-L% were used instead, then A would still be answering B's question yet that response would seemingly not complete the predication.

(1) A: Guess what I got at the store
     B: What did you get?
     A: Oranges (H*L-L%)
     A′: ?Oranges (L*L-L%)

As scripted, the discourse in (1) lacks any mutually believed information that could change how A intends *oranges* to be interpreted, the default H*L-L% is expected to be used. However, in discourse contexts that specify shared knowledge, felicity judgments of these tunes paired appear to shift. Consider two contexts where B brings up additional information assumed to be shared with A. In (2), B brings up a berry allergy, thus creating a context where *blueberries* or *strawberries* would be unexpected responses to the question "What did you get?", but *oranges* would still be ruled in as a possibility; we will refer to this as the ruled-in context. In this context the default intonation H*L-L% is predicted to be felicitous, while L*L-L% seems less felicitous. After A's first turn in (1), the dialogues in (2) and (3) proceed, prompted by the speaker A saying to take a guess.

(2) B: Well, since you know I'm allergic to berries, it can't be blueberries or strawberries... What did you get?
     A: Oranges (H*L-L%)
     A': #Oranges (L*L-L%)

(3) B: Well, since you know I'm allergic to citrus, it can't be lemons or grapefruit... What did you get?
     A: #Oranges (H*L-L%)
     A': Oranges (L*L-L%)

In (3), on the other hand, B brings up a *citrus* allergy as common prior knowledge, which would make *oranges* an unexpected response to the question "What did you get?" Prior to A's unexpected response, at the point in the discourse where B is considering the possibilities of what A has bought, B assumes *oranges* to be mutually known as inappropriate and thus unlikely under a belief that A is a cooperative social agent. In other words, the assumed shared knowledge that B has a citrus allergy

effectively rules out a response of *oranges*, at least from B's perspective. We will refer to this as the ruled-out context. In this context, our subjective impression is that the default H*L-L% is no longer felicitous, while L*L-L% now seems preferred. In both examples, the discourse context establishes shared knowledge between the interlocutors and the response constitutes new information. Prior work would predict that H*L-L% would be a felicitous choice in both contexts, as the default tune for an assertion that signals an update of the shared knowledge. Yet, in the ruled-out context, speaker A has additional knowledge that a response of *oranges* is incongruous with speaker B's expectations based on their shared knowledge of the citrus allergy.

Following [11], it should be more difficult for our speaker (A) to commit to the proposition of $\mathrm{GET}(A, oranges)$ when it is not projected as a possible next step in the discourse. That is, when oranges are ruled in, there is no problem with committing to $\mathrm{GET}(A, oranges)$; when oranges are ruled out, it should be comparably more difficult to make this same commitment. In the ruled-out context, some additional renegotiation of the common ground is needed to accommodate the new proposition.

Based on this characterization of H*L-L% and L*L-L% in the ruled-in and ruled-out contexts, we hypothesize that the felicity of H*L-L% and L*L-L% is dependent on the epistemic stance of the speaker as it relates to a particular discourse context. In particular, H*L-L% is licensed when the asserted proposition is ruled in by the context, and L*L-L% is licensed when the asserted proposition is ruled out by the context. In the following sections, we present two perception experiments designed to elicit participant judgments about whether the proposed relationship between these tunes and contexts holds.

2. Methods

For our experiments, we use short dialogues that provide a conversational context and establish the (assumed) shared knowledge between two conversation partners A and B, like those in (2) and (3). Twelve dialogues, each with a ruled-in and a ruled-out version (24 total items), were recorded by a male and female, native speakers of American English. The final response[1] for each dialogue (e.g,. "oranges") was recorded by the male speaker with naturally produced H*L-L% and L*L-L% intonation. Eight filler dialogues were used with different discourse structures, tunes, and response sentences.

The recording using H*L-L% was used as the source recording for subsequent pitch resynthesis using PSOLA in Praat [12] to generate new H*L-L% and L*L-L% pitch contours with fixed pitch target values based on the male speaker's natural productions of the two tunes as shown in Figure 1. The H*L-L% contour started at 100Hz at the beginning of the accented word and rose to an H* pitch accent with a maximum peak of 140Hz on the stressed syllable before falling to 85Hz at the L- phrasal accent, then finally falling to 80Hz for the L% boundary tone. The L*L-L% contour similarly started at 100Hz before falling to an L* pitch accent with a target of 80Hz on the stressed syllable, which was maintained until the L- tone before finally falling to 75Hz at the L% boundary tone.

We use a 2-alternative forced choice (2AFC) task to elicit felicity judgments from participants regarding the two tunes under investigation within the two contexts we have presented. Each trial displays a conversation presented via text on screen,

---

[1]The complete set of responses are as follows, with the stressed syllable denoted by an acute accent: *óranges, vólleyball, róses, rósemary, ríbeye, nárnia, mónorails, lávender, íowa, gérmany, cinderélla,* and *arúgula.*
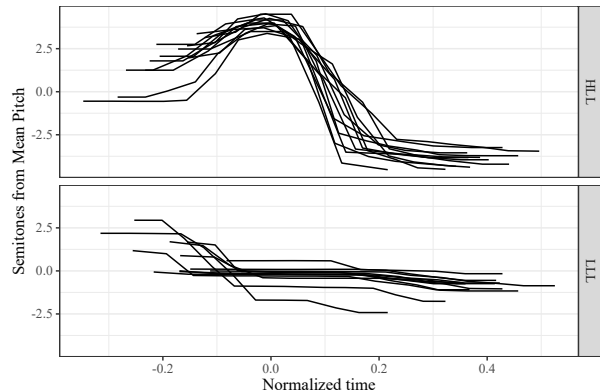

Figure 1: *Nine-point median-smoothed pitch contours with normalized time on the x-axis, centered on H* location at time point 0.0. Contours are given in semitones from the average pitch for each resynthesized contour.*

along with two buttons labeled Choice A and Choice B. At the start of the trial, the participant hears an audio recording of the first four turns in the dialogue. Participants must then click the buttons to listen to the two distinct versions of the dialogue continuation in the male speaker's response. For both the Choice A and Choice B versions, the response word is the same, but the tune used is different. For example, Choice A might be "oranges" with H*L-L% while Choice B is "oranges" with L*L-L%. The text of the dialogue is italicized and shown in all lower case with no punctuation to minimize the influence from a participant's inner reading that may differ from the provided audio.

The assignment of tune to Choice A or B is balanced across trials, resulting in four conditions for each dialogue: 2 contexts (ruled-in, ruled-out) × 2 positions of tunes as Choice A or B. Participants see each dialogue in one context condition in the first block of 12 trials, and the opposite context condition in the second block of 12 trials, with the same order of dialogues and mappings of tunes to Choice A and B in both blocks. Block ordering and the Choice-Tune mapping is counter-balanced in a between-subjects design.

For each trial, participants are tasked with listening to the dialogue and then the two response choices. After listening to both options, we ask participants to "Select the response that you felt was more appropriate given the conversation." Participants are provided with buttons labeled "Select Choice A" and "Select Choice B" to make their selection. For our hypothesis, we predict that the choice(s) mapped to L*L-L% will be more likely to be selected in the ruled-out context while the choice(s) mapped to H*L-L% will be more likely to be selected in the ruled-in context.

Our first experiment used only single-word utterances like "oranges," placing the burden of conveying the speaker's epistemic stance entirely on intonation. Specifically, the use of L*L-L% when expressing some proposition p—*I bought oranges*—may license the conversational implicature of q: the speaker's epistemic stance. In the ruled-out context, q would be equivalent to "I know you're allergic to oranges". The conversational implicature q in the ruled-in context is much less obvious since the speaker should have no reason to believe p does not fit into the discourse context. Of course, in an actual conversation, speakers are typically not as constrained in how they convey their understanding of the discourse context; speakers can

use more than one word to communicate their stance in their response to a question. Our second experiment thus extends the final response in our dialogues from a single-word response such as "oranges" (*p* with H*L-L% or L*L-L%) to include additional propositional content that makes explicit the speaker's epistemic stance *q*. For example, "oranges" is extended to "*I know you like them, so I bought* oranges" in the ruled-in context and "*I know you're allergic, but I bought* oranges" in the ruled-out context.

Including this explication of the speaker's epistemic stance prior to the predication of *oranges*, bearing the nuclear tune, allows us to ask whether intonation's expression of *q* via conversational implicature is made redundant. In particular, we can consider the utterance to contain two cues to the speaker's epistemic stance: the propositional cue from the explication, and an intonational cue from the nuclear tune. The use of two cues leads to two competing hypotheses. First, Hypothesis 1 is that the use of intonation may still be informative even if made redundant, which would predict that the combination of mismatching cues (e.g., H*L-L% in the ruled-out context, where "I know you're allergic" is explicitly stated) will be dispreferred, and there should be a stark difference in the proportion of responses preferring each of the two contours. Alternatively, Hypothesis 2 is that the use of intonation is superfluous in the presence of propositional content explicating the same meaning, predicting that when *q* is made explicit, intonation does not provide any additional information. From Hypothesis 2 we rpedict either that there will be little or no difference in response preference between the two contours, or that the H*L-L% tune will be universally preferred as the default contour for declarative assertions. These hypotheses are evaluated using the same 2AFC procedure as Experiment 1 with new response materials.

For each experiment, we recruited 40 paid participants from Prolific. One participant was removed from each experiment due to self-reporting that they were not a native English speaker. Participants in Experiment 1 were between 18 and 63 years old (mean 33, SD 11); 19 participants self-reported as Female, 17 as Male, and 3 as Other. Participants in Experiment 2 had not participated in Experiment 1 and were between 18 and 52 years old (mean 32, SD 9.9); 23 self-reported as Female and 16 self-reported as Male. In each experiment, 18 participants reported musical experience and 2 reported knowing a second language.

3. Results

Participant behavior is modeled with logistic mixed effects regression with the log-odds of selecting L*L-L% as the dependent variable and fixed effects of context, gender, musical training, knowledge of other languages, block, and counterbalanced stimulus list.The random effects structure for the model includes random intercepts by subject and by dialogue. The stimulus list predictor was sum coded, while all other categorical predictors were scaled-sum coded.[2] Factors with more than 2 levels were Bonferroni-corrected for multiple comparisons. Figure 2 shows the results for Experiment 1 and Experiment 2.

Comparing the aggregated mean proportion of responses for H*L-L% and L*L-L% in each experiment shows a very strong preference for selecting H*L-L% in the ruled-in context, and somewhat less strong preference for selecting L*L-L% in the ruled-out context.

Results from the logistic mixed-effects regression model for Experiment 1 show that the odds of selecting L*L-L% is
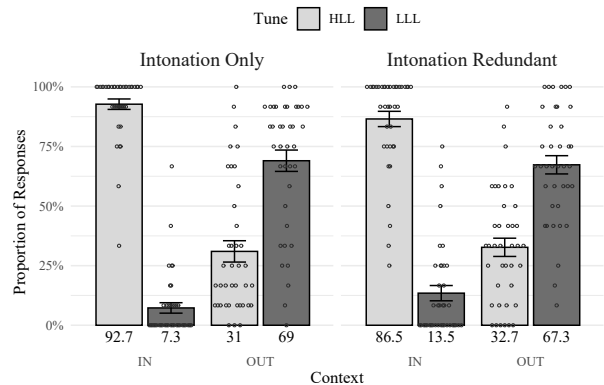
Figure 2: *Results for Experiments 1 (Intonation Only) and 2 (Intonation Redundant). Each dot corresponds to a single participant's response proportion. Bars show mean proportions for data aggregated over all participants with error bars for the ±1 standard error (n=39 for both experiments).*

38 times higher in the ruled-out context than in the ruled-in context ($\hat{\beta} = 3.64$, $z = 16.2$, $p < .001$). We attribute the stronger preference for H*L-L% in the ruled-in context to a possible overall bias for selecting H*L-L%, supporting the view that this is the **default** tune for assertions. This bias is reflected in the intercept for the model, showing lower overall log odds of selecting L*L-L% averaged across both contexts, favoring H*L-L%, does not reach statistical significance ($\hat{\beta}_0 = -0.47$, $z = -1.34$, $p = 0.18$). A significant effect of block ($\hat{\beta} = 0.99$, $z = 5.22$, $p < .001$) appeared, where the odds of selecting L*L-L% in the second block are 2.6 times higher than in the first block. We interpret the effect of block as a correction of the initial H*L-L%-bias yielding roughly equal odds of selecting either tune in the second block when all other predictors are held at their average. Other demographic covariates showed small effects but failed to reach significance.

The aggregate results of Experiment 2 are nearly the same as in Experiment 1: L*L-L% is still strongly preferred in the ruled-out context, but not as much as H*L-L% is preferred in the ruled-in context, despite the additional propositional content. The results of our logistic mixed effects regression model again show a significant effect of context such that the odds of selecting L*L-L% are 18 times higher in the ruled-out context than in the ruled-in context ($\hat{\beta} = 2.92$, $z = 14.8$, $p < .001$). Interestingly, we do not find a significant effect of block in Experiment 2 ($\hat{\beta} = 0.22$, $z = 1.28$, $p = .2$) like we did for Experiment 1. The overall bias for H*L-L% given by the intercept of the model does seem to remain ($\hat{\beta}_0 = -0.62$, $z = -1.72$, $p = .086$), but again fails to reach significance.

The smaller effect of context in Experiment 2 comes as a result of the increased proportion of L*L-L% responses in the ruled-in context (7.3% in Experiment 1 vs. 13.5% in Experiment 2). Further investigation shows that this result is driven by male-identifying participants being more likely than female-identifying participants to select L*L-L% in the ruled-in context. Including an additional interaction between context and gender in the model for Experiment 2 shows that this is statistically significant ($\hat{\beta} = -1.09$, $z = -2.93$, $p = .003$). This interaction was not predicted based on the results of the model for Experiment 1, where including the interaction prevented the

model from converging with random effects. A post-hoc comparison using a simpler model without the random effects structure for Experiment 1 converged but the interaction was not statistically significant ($\hat{\beta} = -.69$, $z = -1.70$, $p = .18$).

4. Discussion

The results from Experiment 1 show that listeners perceive L*L-L% as more appropriate in a discourse context where their response is ruled out by their interlocutor's expectations. In this experiment, intonation served as the only cue to the speaker's epistemic stance, namely, as knowing that their is in some way incongruous with the discourse context, and therefore not a response that their interlocutor would have expected. Under this view, the use of L*L-L% with a sentence expressing the proposition p licenses an inference q about the speaker's knowledge state regarding p: that the speaker knows p is incongruent with the context. Using H*L-L% does not license the inference of q, making the default assertion of p infelicitous.

Experiment 2 further investigated how the pragmatic contribution of intonation may change when potentially made redundant by an alternative cue. We hypothesized that the use of additional propositional content would make intonation's contribution redundant, which may either enhance the contrast between the two contours when the cues mismatched or, alternatively, diminish or neutralize the contrast entirely such that the choice of intonation to cue (or not cue) epistemic stance is rendered unnecessary. However, our results were largely the same as in the first experiment: while H*L-L% was preferred in the ruled-in context, L*L-L% was still preferred in the ruled-out context. This finding speaks against the second hypothesis.

Our experiments used hypernym-hyponym relations to define a response utterance as ruled-in or ruled-out, based on whether its associated proposition was congruent or incongruent with an interlocutor's expectations grounded in shared knowledge. While this account of tune meaning seems satisfactory in capturing the meaning in the question-response contexts scripted in our experimental dialogues, it does not follow that the core pragmatic meaning of L*L-L% is restricted to this very specific meaning, i.e., to encode a mismatch between a referent and its superordinate category. The results of Experiment 2 show that L*L-L% plays a role in utterance interpretation even when epistemic stance is made explicit, hinting at an underlying pragmatic function that in turn gives rise to the epistemic stance-related inference in our experimental dialogue. What could this more general meaning for L*L-L% be?

We propose that the function of the L*L-L% tune on a sentence expressing a proposition p is to **convey the speaker's withheld commitment to** p. As such, the speaker effectively introduces p to the discourse without sanctioning the addition of p to the common ground. The function of withholding commitment is not the same as a denial of commitment to p and also contrasts with the function of H*L-L% which, when paired with a sentence expressing p, expresses a commitment to p that is not contingent [2] on their interlocutor's subsequent contribution. The interlocutor is thus licensed to hold the speaker accountable for p [13]. In this vein, [14] empirically showed that participants were more likely to punish speakers who overcommit to unreliable information over speakers who took strategies to reduce their commitment to the asserted information.

Considering our experiment in relation to [14], our speaker in the ruled-out context should realize that the new information to be asserted, for which the H*L-L% tune is the default, will not be taken well by their interlocutor given their known citrus

allergy. Note that the speaker does not have the option to simply "call off" their commitment [15] through the use of rising intonation (i.e., L*H-H%), which puts the onus of commitment on the interlocutor, since the interlocutor's question of "What did you buy?" prompted the problematic response. The use of L*L-L% over H*L-L% is then marked, inviting the question of why the speaker is not overtly committing themselves to their stated proposition yet also not placing contingent commitment elsewhere. In our face-threatening ruled-out context, the speaker may use the marked L*L-L% tune to signal acknowledgment that the propositional content of their utterance is not expected and will not be taken well by their interlocutor.

If the choice of tune in our experiment serves as a cue to speaker commitment, rather than a direct cue to epistemic stance (i.e., acknowledging awareness of the discourse context in our experiment), it follows that including an explicit statement of the speaker's epistemic stance, as in Experiment 2, does not render the use of L*L-L% redundant. The L*L-L% tune makes an independent meaning contribution beyond epistemic meaning. Viewed in this way, we are not surprised to see essentially the same results between the two experiments. The additional propositional content q in the extended sentences from Experiment 2 merely reinforce the conversational implicature licensed from the speaker's withholding of commitment to p. Even though the speaker still ultimately becomes committed to p, working out the implicature licensed from the use of L*L-L% may help to lessen the imminent reputational damage.

5. Conclusions

In this work we provided two related experiments to empirically investigate the pragmatic meaning of the L*L-L% tune in American English as it relates to epistemic stance. We found through our first experiment that L*L-L% with a declarative sentence is interpreted as more felicitous than the default H*L-L% in contexts where a speaker's response has been ruled out by their interlocutor's expectations based on knowledge assumed to be shared. Curiously, the same pattern of results emerged even when we tried to make intonation's contribution redundant through an additional cue using propositional content to explicate the speaker's epistemic stance. Combined, these results hint at the need for a more general characterization of the meaning of L*L-L%, from which the speaker's epistemic stance can be inferred without being directly encoded by the intonational tune. We proposed that L*L-L% may be described as a withholding of commitment to an asserted proposition p, licensing inferences about why the speaker is not being forthcoming.

While a complete semantic and pragmatic account of "withholding" is beyond the scope of this paper, work such as the commitment space semantics model proposed by [11] would likely be a frutiful way forward. Moreover, this work has made new steps towards describing the contrast between H*L-L% and L*L-L%, which has been sorely lacking despite being a minimal pair differing in only the pitch accent. Lastly, while we did not seek to test the difference between H* and L* *specifically* nor whether the meaning of L*L-L% is compositional, this work provides a step towards understanding these issues.

6. Acknowledgments

7. References

[1] J. Pierrehumbert and J. Hirschberg, *The meaning of intonational contours in the interpretation of discourse*. Cambridge Massachusetts: MIT Press, 01 1990, ch. 14.

[2] C. Gunlogson, "A question of commitment," *Belgian Journal of Linguistics*, vol. 22, pp. 101–136, 12 2008.

[3] R. S. Burdin and J. Tyler, "Rises inform, and plateaus remind: Exploring the epistemic meanings of "list intonation" in american english," *Journal of Pragmatics*, vol. 136, pp. 97–114, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037821661730663X

[4] C. Bartels, *Towards a compositional interpretation of English statement and question intonation*. University of Massachusetts Amherst, 1997.

[5] D. Büring, *Intonation and meaning*. Oxford University Press, 2016.

[6] S. Jeong, "Intonation and sentence type conventions: Two types of rising declaratives," *J. Semant.*, vol. 35, pp. 305–356, 2018.

[7] M. Beckman, J. Hirschberg, and S. Shattuck-Hufnagel, *The Original ToBi System and the Evolution of the ToBi Framework*, 01 2004, pp. 9–54.

[8] C. Gussenhoven, *The Phonology of Tone and Intonation (Cambridge University Press, 2004, reprint with corrections 2008).*, 07 2004.

[9] D. F. Farkas and K. B. Bruce, "On Reacting to Assertions and Polar Questions," *Journal of Semantics*, vol. 27, no. 1, pp. 81–118, 2 2010. [Online]. Available: https://academic.oup.com/jos/article-lookup/doi/10.1093/jos/ffp010

[10] J. R. Hobbs, "The Pierrehumbert-Hirschberg Theory of Intonational Meaning Made Simple: Comments on Pierrehumbert and Hirschberg," in *Intentions in communication*, P. Cohen, J. Morgan, and M. Pollack, Eds., 1990, ch. 15, pp. 313–323.

[11] M. Krifka, "Bias in commitment space semantics: Declarative questions, negated quetions, and question tags," in *Semantics and Linguistic Theory*, vol. 25, 2015, pp. 328–345.

[12] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [computer program]. version 6.1.27. retrieved june 2, 2020," 2020.

[13] B. Geurts, "Communication as commitment sharing: speech acts, implicatures, common ground," *Theoretical Linguistics*, vol. 45, no. 1-2, pp. 1–30, 2019.

[14] D. Mazzarella, R. Reinecke, I. Noveck, and H. Mercier, "Saying, presupposing and implicating: How pragmatics modulates commitment," *Journal of Pragmatics*, vol. 133, 08 2018.

[15] D. Rudin, "Rising above commitment," Ph.D. dissertation, UC Santa Cruz, 2018.