

# Polarity Reversal Questions and the Semantics of Prosodic Incorporation<sup>1</sup>

Tom ROBERTS — *Utrecht University*

Deniz RUDIN — *University of Southern California*

**Abstract.** English polarity reversal questions (PRQs, e.g. *Gertrude mowed the lawn, didn't she?*) comprise individual biased questions despite the fact they consist of two apparently independent clauses, which, if uttered as two prosodically distinct sentences (*Gertrude mowed the lawn. Didn't she?*) give instead the sense that the speaker is backtracking on their claim. Many recent accounts stipulate the special discourse effects of PRQs into their context update potential at the level of a construction (Malamud and Stephenson, 2015; Farkas and Roelofsen, 2017; Bill and Koev, 2023: a.o.). We propose that the range of interpretations of PRQs, and their difference from string-identical sequences of two sentences, can be attributed to the fact that PRQs are prosodically integrated, packaging two syntactically independent clauses into a single Intonational Phrase and thus a single context update. We argue that this assumption, combined with a vanilla treatment of the discourse effects of uttering declarative and interrogative clauses, can derive the interpretations of PRQs, explain important limitations on their form, and account for their differences from non-integrated sequences, without relying on construction-specific stipulations.

**Keywords:** tag questions, polarity reversal questions, context update, prosody, semantics-pragmatics interface, Table model

## 1. Introduction

The form of a sentence relates in principled ways to the discourse effects of uttering that sentence. For instance, declaratives canonically express **assertions**—supplying information—whereas interrogatives canonically express **questions**—requesting information. This robust correlation between form and interpretation might suggest a straightforward mapping from the former to the latter: that the form of the sentence in some way *encodes* its illocutionary force.

However, the world is not so simple. Syntactically identical sentences can be used to different conversational ends, such as rhetorical questions, which seem to comprise assertions despite interrogative syntax (Han, 2002; Caponigro and Sprouse, 2007; Biezma and Rawlins, 2017: a.o.), and reportative evidentials, which are not always assertive despite their declarative syntax (AnderBois, 2014; Faller, 2019; Pancheva and Rudin, 2019: a.o). Moreover, similar illocutionary forces can be expressed using sentences of different syntactic types: both interrogative and declarative sentences can be used to ask questions if accompanied by final rising intonation (Hirschberg and Ward, 1995; Gunlogson, 2001, 2008; Malamud and Stephenson, 2015; Farkas and Roelofsen, 2017; Jeong, 2018; Westera, 2018; Rudin, 2022: a.o.).

So while the syntactic form of a sentence is generally a reliable cue for the kind of effect its utterance will have in discourse, we must take into account factors like intonation and discourse context before we can derive a complete account of the discourse effect of an utterance. We need models of how clause typing affects context-update potential, but also models of how other

---

<sup>1</sup>Thanks to Samir Alam, Karl DeVries, Donka Farkas, Beste Kamali, Travis Major, Floris Roelofsen, Paul Willis, Danfeng Wu, and three anonymous reviewers for helpful comments and suggestions. All errors are our home-brewed contribution. This research was partially supported by NWO grant 406.18.TW.009 “A Sentence Uttered Makes a World Appear—Natural Language Interpretation as Abductive Model Generation” awarded to Muskens.

factors such as intonation modulate that context-update potential, and of how general rules of *pragmatics* derive additional inferences downstream from the conventionalized context-update potential of an utterance (Roberts, 2012; Farkas and Roelofsen, 2017; Murray and Starr, 2018; Roberts, 2023b: a.m.o.)

This paper considers the interaction of conventionalized context-update potential and general rules of pragmatic inference via a case study of English **Polarity Reversal Questions** (PRQs), as in (1):

- (1) a. You sent the envelope, didn't you?
- b. You didn't send the envelope, did you?

PRQs are chimerical: syntactically, they appear to be a sequence of two separate root clauses, a root declarative clause followed by a root interrogative clause of opposite polarity that has undergone VPE (Sailor 2009, 2014). But their discourse effect is, intuitively, to ask a single (biased) question, roughly a request for confirmation of what the speaker is tentatively committed to (Sadock, 1971; Ladd, 1981: a.m.o.). Given the appearance of bi-clausality and the assumption that interpretation follows form, we want to know: why are PRQs interpreted the way they are, in view of the form they have?

An initial reasonable proposal might look something like this. PRQs look like a declarative followed by an interrogative, so we could simply treat them as such (see e.g. Asher and Reese 2007). But there is a wrinkle to the reasonable proposal: string-identical **Non-Integrated Sequences** (NISs) of a declarative followed by an interrogative produce a different discourse effect to PRQs:

- (2) a. You sent the envelope. #Didn't you (send the envelope)?
- b. You didn't send the envelope. #Did you (send the envelope)?

Whereas the sequences in (1) comprise individual biased questions, those in (2) signify internal conflict: that the speaker is 'backtracking' on their assertion, and deciding to ask the question instead. This contrast has been used as evidence against treating PRQs as simple declarative + interrogative sequences (Northrup 2014), motivating analyses capturing their discourse effects with *sui generis* stipulations. For instance, Krifka (2015) analyzes PRQs using a covert speech-act level disjunction operator; Malamud and Stephenson (2015) assign them a non-compositional discourse effect of projected speaker commitment to the denotation of the declarative; Farkas and Roelofsen (2017) assign them an additional 'marked' discourse effect of signalling the speaker's evidence-based credence toward the denotation of the declarative, depending on intonation; Bill and Koev (2023) relate the declarative to the question via a covert TAG operator that drapes an epistemic possibility modal over the denotation of the declarative. A recent exception is Scheffler and Malamud (2023), who present a compositional analysis of a subset of PRQs ending in *won't you*. Though their account is compositional, they nonetheless treat the interrogative tag as altering the discourse effect of the declarative, lowering the speaker's degree of commitment to the proposition it denotes. Every account of the phenomenon of which we are aware shares the feature that lack of full speaker commitment of the kind ordinarily associated with assertions of declarative sentences is hard-coded into the conventional discourse effect of PRQs, making them something other than a simple sequence of an ordinary assertion followed by an ordinary polar question.

## Polarity Reversal Questions

In this paper we argue that there is hope for the reasonable proposal, leveraging a neglected fact about the difference in form between PRQs and NIS: **prosody**. We propose that both PRQs and NISs are, in fact, what they look like: a declarative sentence followed by an interrogative sentence, cashed out on the level of context update as an ordinary assertive update potential and an ordinary question update potential. They differ in that PRQs, and not NISs, are prosodically integrated into a single Intonational Phrase (IP), which we take to individuate context updates.

We show that this simple assumption, in combination with a general theory of context update and a general theory of pragmatics, can not only derive the right interpretations for PRQs and their difference from NISs, but also explain why PRQs necessarily involve polarity reversal. NISs comprise an assertion, which projects only one future for the conversation, with a polar question, which projects two, resulting in pragmatic incoherence. But PRQs package an assertion and a polar question into a single composed context update, ‘skipping’ to a context in which the speaker is committed to  $p$  but is projecting both  $p$  and  $\neg p$  as possible future additions to the common ground. The incoherence of projecting only one possible future for the conversation before subsequently projecting a second one is avoided; the tension between the speaker’s commitment to  $p$  and projection of  $\neg p$  as a possible future addition to the common ground drives pragmatic inferences that derive the PRQ’s characteristic bias. PRQs are, in fact, what they look like: a sequence of a declarative sentence, with an ordinary assertive update potential, and an interrogative sentence, with an ordinary polar question update potential, prosodically packaged into a single composed update.

### 2. Empirical Landscape

In this section we lay out the empirical ground we aim to cover. First, let us be clear about what, exactly, a polarity reversal question is. Morphosyntactically, PRQs have the anatomy of two separate root clauses: a declarative **anchor**  $p$  followed by an interrogative **tag** of opposite polarity  $\neg p$ ? These are exemplified with positive anchors in (3) and negative anchors in (4).

- (3) a. Sarah can suplex a refrigerator, can’t she?  
b. It will rain tomorrow, won’t it?  
c. You turned off the stove, didn’t you?
- (4) a. Sarah can’t suplex a refrigerator, can she?  
b. It won’t rain tomorrow, will it?  
c. You didn’t turn off the stove, did you?

In addition to the polarity reversal, the tag must also have the same T as the anchor (5), and an elided verb phrase (6) (Sailor, 2009, 2014).

- (5) \*Sarah can suplex a refrigerator, won’t she?
- (6) \*Sarah can suplex a refrigerator, can’t she suplex a refrigerator?

A negative tag can have both preposed (‘high’) or non-preposed (‘low’) negation.

- (7) It will rain tomorrow, won’t it/will it not?

Having established what a PRQ is, we now lay out our analytical desiderata. The first item on the agenda is interpretation. PRQs contribute one of two kinds of speech acts, depending on the context of utterance.

The first species we will term *confirmation questions*: The speaker tentatively asserts  $p$ , albeit not with full confidence, and asks for confirmation of whether  $p$ .

- (8) [Context: the speaker thought that their addressee was in Bucharest, but they appear at the speaker's doorstep in Pasadena]  
You were going to fly to Bucharest, weren't you?

In this context, the speaker is understood to be second-guessing their commitment about the addressee's plans. Depending on how the addressee responds, the speaker may rescind their commitment about what the addressee's plans were.

The second kind of PRQs we call *pretense-questions*: The speaker makes the pretense of asking whether  $p$ , though it is clear they are committed to  $p$ .

- (9) [Context: The addressee has just asked the speaker why they're watching a 40-minute youtube video about the battle of Midway]  
I'm a 40-year-old man, aren't I?

In this context, the speaker is not understood to be second-guessing their own age, but implying that the answer to the previous question should've been obvious.

Ideally, these different discourse effects would be derived from the interaction between discourse contexts and the semantics/pragmatics of PRQs.

### 2.1. Puzzle 1: Non-integrated sequences

As we have seen, non-integrated sequences are stringwise equivalent to PRQs, but prosodically and interpretively distinct. The intuition is that NISs involve a 'pause' in between declarative and interrogative, and have the marked interpretation that the speaker is second-guessing the commitment they just made. This gives the sense that the speaker is somehow retracting, or at least backing off from, their immediately prior claim. We represent the pragmatic oddness of this maneuver with #.

- (10) a. Sarah can suplex a refrigerator. #Can't she?  
b. You didn't turn off the stove. #Did you?

### 2.2. Puzzle 2: Polarity reversal

A hallmark of PRQs is PR. This turns out to be significant, because same-polarity tag questions exhibit sharply different effects from PRQs. Consider, for instance, a biclausal tag question in which the tag and anchor both have positive polarity (a 'positive-positive' question):

- (11) You sent the envelope, did you?

Such questions are natural, but convey a meaning unlike either of the flavors of PRQs we have seen. Whereas PRQs generally involve some sense of commitment on the part of the speaker to the truth of the anchor  $p$ , positive-positive tag Qs convey roughly that the speaker is incredulous or skeptical that  $p$  is the case. For instance, (11) is naturally uttered in a context where the addressee claims they sent the envelope, but the speaker thinks they are lying.

## Polarity Reversal Questions

Negative-negative tag Qs are, by stark contrast, sharply unacceptable, and virtually absent from English corpora (see e.g. Tottie and Hoffmann 2006).

(12) \*You didn't send the envelope, didn't you?

Although many semantic theories of questions assign identical denotations to positive and negative questions (e.g. Hamblin 1973; Groenendijk and Stokhof 1984), the divergent behavior of tag questions depending on the relation of the polarity of the anchor and tag suggests that the differences in context update potential between positive and negative questions plays an important role in any account of the discourse effects of PRQs.

### 2.3. Summary and analytical desiderata

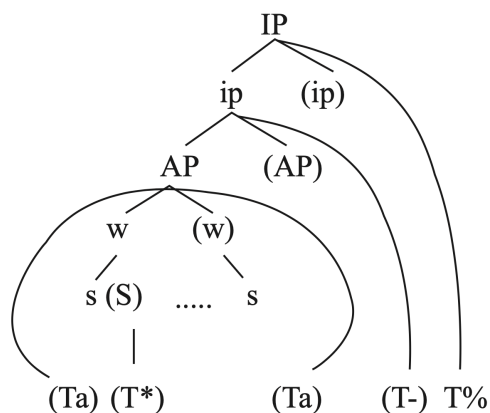
We aim to account for three aspects of PRQ meaning in this paper:

1. The range of discourse effects of PRQs, ideally **derived** from their form, general principles of context update, and general principles of pragmatic interpretation
2. The NIS puzzle: NISs seem to indicate speaker changing their mind across two updates, but PRQs seem to comprise a single coherent update
3. The polarity reversal puzzle: why positive same-polarity tags are possible but interpretively distinct from PRQs, and why negative same-polarity tags are simply impossible

### 3. Form: Prosody

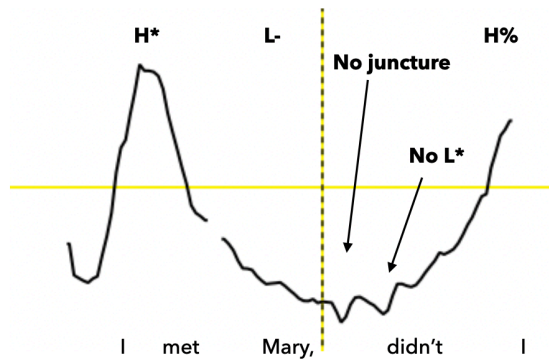
The crucial characteristic of PRQs distinguishing them from NISs is their prosody. We make the following assumptions about prosodic phonology (see Jun 2022 for an excellent overview). The maximal prosodic unit is the Intonational Phrase (IP). IPs delineate (prosodically) independent utterances and host terminal contours: a nuclear pitch accent (\*), followed by a phrase accent (-) and boundary tone (%), plus optional leading or trailing tones.

(13) Intonation and the Prosodic Hierarchy (Jun 2022 ex.1)



As a consequence, IPs can be empirically identified both by intonation (presence of a terminal contour) and by prosody (separated from other IPs by a large juncture—Jun 2022 ex.2). By both diagnostics, integrated PRQs occupy a single IP:

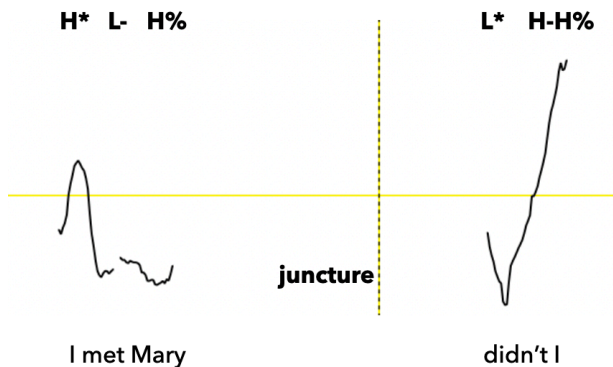
(14) Pitch track of *I met Mary, didn't I?*



PRQ intonation falls over the declarative anchor and rises over the interrogative tag, just like in ordinary independent falling declaratives and rising interrogatives. However, in integrated PRQs, there is just a single terminal contour: there is only an H\* pitch accent in the first clause, and no L\* pitch accent in second clause, which would be realized as a significant local minimum in pitch. In addition, there is no sizeable juncture between clauses. Both diagnostics show that PRQs are prosodically “packaged” into a single IP. **PRQs comprise two sentences syntactically, packaged into one utterance prosodically.**

In the case of non-integrated sequences, we see the opposite; each sentence occupies a separate IP:

(15) Pitch track of *I met Mary. Didn't I?*



In addition to the obvious juncture separating the two clauses, we see evidence for two separate terminal contours: there is a L\* pitch accent in the second clause, realized as a significant local minimum in pitch, in addition to the H\* pitch accent in the first clause. NISs show a more standard relation between syntax and prosody: two sentences syntactically, packaged into two utterances prosodically.

**4. Analysis step 1: IP as domain of update**

Prior work on conventional discourse effects has sometimes (implicitly) assumed that the IP is the domain of context update. For instance, Farkas and Roelofsen (2017) propose that context update is enacted by an UTTERANCE function; in prosodic phonology, the IP is the prosodic domain of an independent utterance. And many works (e.g. Gunlogson 2001, 2008; Malamud and Stephenson 2015; Jeong 2018; Rudin 2022) have analyzed intonational tunes as modifying

## Polarity Reversal Questions

context update potential; this presupposes that context updates are packaged into IPs, as IPs are the hosts of the terminal contours to which these works attribute such effects. We make this assumption explicit: **the prosodic unit of the IP individuates context updates**. In other words, if you’ve got just one IP, then you’ve got just one context update.

We assume an utterance function UTT that maps sentences to their update potentials (Farkas and Roelofsen 2017), i.e. functions from contexts to contexts. It applies to IPs. Because IPs ordinarily host a single sentence, this will ordinarily work exactly the same as models on which context-update functions are applied to sentences, not to IPs. But we observed in the previous section that PRQs prosodically package two sentences into a single IP. So what happens if two syntactically distinct sentences are packaged into a single context update? A simple first-pass analysis: update function composition.

(16)  $UTT(IP) = UTT(S_1) \circ \dots \circ UTT(S_n)$ , where  $S_1, \dots, S_n$  is the sequence of sentences contained within IP<sup>2</sup>

UTT returns a single update potential: the result of composing the update potentials it assigns to each sentence in the ordered sequence of sentences within the IP. Calling this a “simple first-pass analysis” might pragmatically implicate that it is too simple, and must be tweaked. But in what follows, we argue that it does the job perfectly well.

### 5. Implementation: Table

To make our assumptions about the mechanisms of context update explicit, we couch our analysis in the Table model of discourse (Farkas and Bruce 2010). A context in the Table model is typically defined using four components, though two are primitive and two are derived:

- (17) THE TABLE MODEL: a context  $c = \langle DS_c, CG_c, T_c, PS_c \rangle$
- a. INDIVIDUAL DISCOURSE COMMITMENTS  
 $DS_c$  is the set of commitment states  $DC_{X,c}$  for each interlocutor  $X$  in  $c$   
For any agent  $X$ ,  $DC_X$  is the set of propositions  $X$  is committed to in  $c$
  - b. THE COMMON GROUND  
 $CG_c$  is the set of all propositions that all interlocutors are committed to  
( $= \bigcap DC_c$ )
  - c. THE TABLE  
 $T_c$  is a stack of “Issues” (sets of propositions), the maximal element of which ( $MAX(T_c)$ ) is the immediate Question Under Discussion (QUD) in  $c$  (Roberts 2012; Ginzburg 1996)
  - d. THE PROJECTED SET  
 $PS_c$  is the set of all future Common Grounds that could result from adding an element of ( $MAX(T_c)$ ) to  $CG_c$   
( $= \{ CG_c + p : p \in MAX(T_c) \}$ )

Context update potentials for utterances are defined in terms of how they update the speaker’s discourse commitments, and how they update the Table. Following Farkas and Bruce (2010), we assume that ordinary assertions (expressed by uttering a falling declarative  $p$ ) add  $p$  to the speaker’s discourse commitments and push a singleton QUD whose only resolution is  $p$  onto

---

<sup>2</sup>We define UTT( $S$ ) for the relevant sentence types in (18) below.

the Table (18). Note that this addition to the Table also results in a singleton projected set.<sup>3</sup>

(18) Where  $\llbracket He ate the cake \rrbracket = p$

$$a. \text{UTT}(He ate the cake) = \lambda c. \left[ \begin{array}{l} DC_{sp} = DC_{sp,c} + p \\ T = T_c + \{p\} \\ PS = \{CG + p\} \\ c' = c \text{ in all other respects} \end{array} \right]^{c'}$$

Ordinary polar questions (expressed by uttering a rising interrogative  $p$ ?) add nothing to the speaker's discourse commitments, but push the QUD  $\{p, \neg p\}$  onto the Table (19).

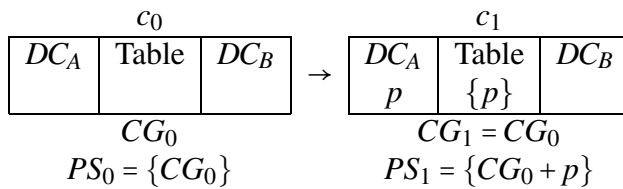
$$(19) \text{UTT}(Did he eat the cake?) = \lambda c. \left[ \begin{array}{l} T = T_c + \{p, \neg p\} \\ PS = \{CG + p, CG + \neg p\} \\ c' = c \text{ in all other respects} \end{array} \right]^{c'}$$

With these definitions in hand, we can now unpack the differences in context update between PRQs and NISs.

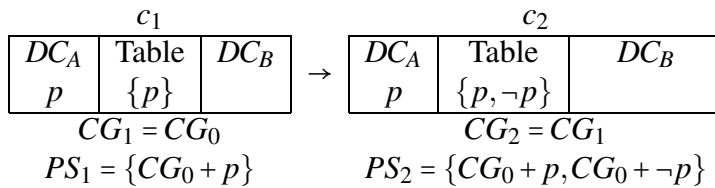
### 5.1. NISs and PRQs in the Table

NISs occupy separate IPs, and so will comprise separate updates. Specifically, the two-step sequence of an ordinary assertion followed by an ordinary polar question:<sup>4</sup>

(20) a. Utterance of *He ate the cake.* (=  $p$ )



b. Subsequent utterance of *Didn't he (eat the cake)?*



Integrated PRQs occupy a single IP, and so will comprise a single update, derived by composing the updates either sentence would instantiate on its own into a single step:

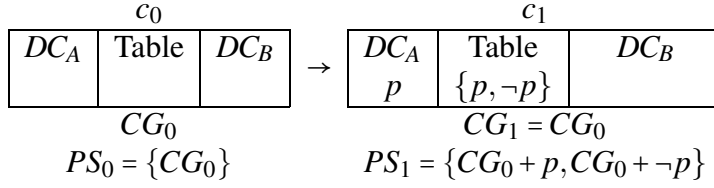
(21) Utterance of *He ate the cake* ◦ utterance of *Didn't he (eat the cake)?*  
= Utterance of *He ate the cake, didn't he?*

<sup>3</sup>Here we simply stipulate these update potentials, but they can be derived from a general definition of UTT + general assumptions about the semantics of declarative and interrogative clauses & general assumptions about the contributions of rising and falling intonation; see Farkas and Roelofsen (2017); Rudin (2022) for details. We do not analyze the semantic contribution of intonational tunes here.

<sup>4</sup>Though the Table is defined as a stack, we represent only the maximal element of the stack in these diagrams, as the presence of  $\{p\}$  beneath  $\{p, \neg p\}$  on the stack plays no role in our account. Any resolution of  $\{p, \neg p\}$  will also resolve  $\{p\}$ , so its presence is immaterial.



## Polarity Reversal Questions



While (20) and (21) result in identical output contexts, they differ in the steps by which they arrive there. NISs represent an intermediary stage of the context in which the speaker has projected only  $CG + p$ , but PRQs represent no such intermediary stage. As we will see, given independently motivated assumptions about how comprehenders reason about speakers' doxastic states given the context update they've made, these two paths result in markedly different pragmatic effects in spite of yielding the same output context.

### 6. Analysis step 2: Table pragmatics

On the analysis pursued so far, context update potential is conventionalized. It's determined by the form of an utterance—syntactic clause typing, prosodic packaging, and so on. But not *everything* about discourse is conventionalized. Just as in truth-conditional semantics, we have to discriminate between the conventionalized and the pragmatic in the domain of context update. Figuring out the holistic effect of an utterance requires understanding both what the conventional discourse effect of an utterance is, and what additional inferences are pragmatically licensed by observation that a speaker has updated the context using that discourse effect (Gunlogson 2001; Murray and Starr 2020; Rudin 2022; Scheffler and Malamud 2023 a.m.o.).

In the Table model, context updates involve incurring (or avoiding) speaker commitments, and including (or withholding) propositions within what is placed on the Table. To understand the overall space of inferences that an utterance gives rise to, we need to articulate the pragmatic principles governing the cooperativity of making such moves.

#### 6.1. Maxims for commitment and projection

Here we assume Rudin's (2018; 2022) formalization of the underlying pragmatics of the Table model. Commitment-making is subject to a maxim<sup>5</sup> of SINCERITY ( $\approx$  Gricean QUALITY):

- (22) SINCERITY:  
Violated by any move that adds  $p$  to  $DC_{sp}$ , where  $DOX_{sp} \subseteq p$

This maxim states that a cooperative agent  $X$ 's public commitments ( $DC_X$ ) should accurately reflect what they really believe ( $DOX_X$ ).

Projecting CGs is subject to a maxim of VIABILITY:

- (23) VIABILITY:  
Violated by any move that adds a set including  $p$  to  $T$ , where  $\cap DC_X \cap p = \emptyset$  for some interlocutor  $X$ , or  $DOX_X \cap p = \emptyset$  for some interlocutor  $X$

Adding propositions to the Table projects possible futures for the conversation in which that proposition has become common ground. This maxim states that a cooperative agent shouldn't

<sup>5</sup>Rudin (2022) also discusses a maxim of PUBLICITY that obligates speakers to make relevant commitments if they can; this maxim won't be relevant to our purposes here.

project a possible future for the conversation that contradicts any interlocutor’s commitments (or private beliefs), as that possible future state of the common ground would either be unreachable, as common ground is blocked by the incompatible commitment, or uncooperative, as it would not accurately reflect the beliefs of the interlocutors.

A corresponding pragmatic pressure, COMPREHENSIVENESS, compels agents to project as inclusive as possible a set of paths forward for the common ground, modulo VIABILITY:

- (24) COMPREHENSIVENESS:  
 Violated by any move that adds a set  $P$  to  $T$  such that  $\cup P$  does not include at least one world  $w$  such that  $w \in \cap CG$  and  $CG + \{w\}$  is viable

The set of propositions placed on the Table gives the set of possible future states of the common ground in which the QUD it raises has been successfully resolved. This maxim states that a cooperative agent shouldn’t exclude worlds compatible with the common ground from the space of possible future states from the common ground unless those worlds are unviable.

## 6.2. Application to NISs

Let’s return to the non-integrated update sequence from above:

- (25) a. Utterance of *He ate the cake.* ( $= p$ )
- |                   |       |        |   |                       |         |        |
|-------------------|-------|--------|---|-----------------------|---------|--------|
| $c_0$             |       |        | → | $c_1$                 |         |        |
| $DC_A$            | Table | $DC_B$ |   | $DC_A$                | Table   | $DC_B$ |
|                   |       |        |   | $p$                   | $\{p\}$ |        |
| $CG_0$            |       |        |   | $CG_1 = CG_0$         |         |        |
| $PS_0 = \{CG_0\}$ |       |        |   | $PS_1 = \{CG_0 + p\}$ |         |        |
- b. Subsequent utterance of *Didn’t he (eat the cake)?*
- |                       |         |        |   |                                      |                 |        |
|-----------------------|---------|--------|---|--------------------------------------|-----------------|--------|
| $c_1$                 |         |        | → | $c_2$                                |                 |        |
| $DC_A$                | Table   | $DC_B$ |   | $DC_A$                               | Table           | $DC_B$ |
| $p$                   | $\{p\}$ |        |   | $p$                                  | $\{p, \neg p\}$ |        |
| $CG_1 = CG_0$         |         |        |   | $CG_2 = CG_1$                        |                 |        |
| $PS_1 = \{CG_0 + p\}$ |         |        |   | $PS_2 = \{CG_0 + p, CG_0 + \neg p\}$ |                 |        |

In the first step of this sequence, the speaker moves to a state of the context in which the singleton set  $\{p\}$  is on the Table, projecting only one possible future state of the conversation: one in which  $p$  has become Common Ground. In making this move, the speaker has excluded paths forward in which  $\neg p$  becomes Common Ground. By COMPREHENSIVENESS, this is cooperative only if the speaker believes that  $\neg p$  is not VIABLE.

In the second step of this sequence, the speaker subsequently moves to a state of the context in which the dual set  $\{p, \neg p\}$  is on the Table, projecting two possible future states of the conversation: one in which  $p$  has become Common Ground and one in which  $\neg p$  has. This is only cooperative if the speaker believes that  $\neg p$  is VIABLE.

The speaker’s projection behavior suggests that they take  $\neg p$  to be viable, and that they take  $\neg p$  to be unviable: the two moves **express contradictory demands** on the maxim of VIABILITY. If the speaker thinks  $\neg p$  is viable, then their first update is uncooperative. If the speaker thinks  $\neg p$  is unviable, then their second update is uncooperative. There is no coherent way to reconcile

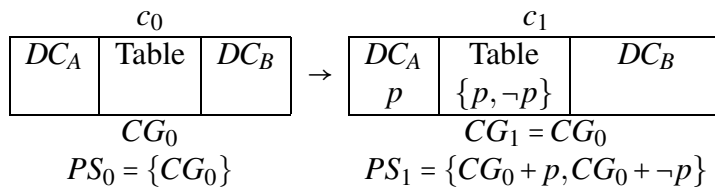
## Polarity Reversal Questions

both of the speaker’s moves with the pragmatics of projection. Hence, the interpretation that the speaker has changed their mind and is backing off from the initial utterance: the only way to understand this update sequence as cooperative is if the speaker has changed their mind about the viability of  $\neg p$  between the two updates.

### 6.3. Application to PRQs

Now let’s consider the composed update, repeated here:

- (26) Utterance of *He ate the cake*  $\circ$  utterance of *Didn’t he (eat the cake)?*  
 = Utterance of *He ate the cake, didn’t he?*



In the composed update, the speaker never puts the context into a state projecting only  $CG + p$ . This means that, unlike in the NIS case above, the speaker never makes a move that is cooperative only if they take  $\neg p$  to be unviable. So the fact that they’ve projected  $\neg p$  does not result in the pragmatic incoherence of the non-integrated sequence, as contradictory demands have not been placed on the maxim of VIABILITY.

There is, however, a different sort of pragmatic phenomenon going on here. Though the speaker has not placed contradictory demands on the maxim of VIABILITY, they have committed an apparent violation of it: their own commitment to  $p$  renders the projected  $CG + \neg p$  unviable. Here, rather than a sequence of moves that impose contradictory demands on one and the same maxim, we have a classic instance of clashing maxims: the speaker has set two maxims in tension with each other in order to productively generate inferences.

On the one hand, given SINCERITY, the speaker’s commitment to  $p$  suggests that they are sure that  $p$  is true. On the other, given VIABILITY, the speaker’s projection of  $CG + \neg p$  suggests that they believe a shared commitment to  $\neg p$  could be cooperatively reached. This tension could be resolved in one of two ways. Either the speaker is understood to signal less-than-complete confidence in  $p$ —their commitment isn’t entirely SINCERE, in the technical sense, and could be walked back (q.v. Scheffler and Malamud 2023); or the speaker is understood to signal that there is only the PRETENSE of a question being asked, and the only acceptable answer is  $p$ —the projected  $CG + \neg p$  is not VIABLE.

In other words, the pragmatics of PRQs can be resolved into either of the two categories outlined above:<sup>6</sup>

- (27) **Legitimate question, less-than-full commitment:**  
 [Context: the speaker thought that their addressee was in Bucharest, but they appear at

---

<sup>6</sup>Note the similarity between these two interpretations and the “inner” and “outer” readings of high-negation polar questions discussed in the literature (Ladd 1981). This parallel may not be accidental: high-negation polar questions likewise involve a tension between asking a question and signalling a speaker’s bias about the answer, which is likely highly conventionalized (Goodhue 2022).

the speaker's doorstep in Pasadena]

You're were going to fly to Bucharest, weren't you?

(28) **Pretense-question, full commitment:**

[Context: The addressee has just asked the speaker why they're watching a 40-minute youtube video about the battle of Midway]

I'm a 40-year-old man, aren't I?

Full commitment readings pose a challenge for accounts of PRQs that hard-code deficient speaker commitment into their update potentials, a category into which the entire prior literature of which we're aware falls (see discussion in introduction). But these readings fall out as one possible resolution of the SINCERITY/VIABILITY clash if bias is derived pragmatically, not conventionally encoded. The contextual variability of the bias of PRQs supports a view on which at least some kinds of question bias are not fully conventionalized, but rather are the result of inferences derived from pragmatic reasoning about the non-canonical state into which a 'biased question' puts the discourse.

### 7. Analysis step 3: The polarity puzzle

The last remaining puzzle concerns the interpretation of same-polarity tags. In popular frameworks of question meaning like Alternative Semantics (Hamblin, 1973) and Partition Semantics (Groenendijk and Stokhof, 1984), positive and negative versions of the same polar question are denotationally equivalent. This assumption might lead us to expect tag questions to be interpreted identically regardless of tag polarity, but as we saw, this is not the case. We propose the polarity restriction of PRQs can be explained by the interaction of uttering positive vs. negative anchors with independently-required felicity requirements of negative polar questions. In what follows we remain agnostic about where these additional restrictions on negative polar questions come from.

#### 7.1. Negative-negative tags

Recall from §2 that negative-negative tag Qs are outright incoherent, regardless of the position of negation in the tag (\* here indicates systematic unacceptability, rather than ungrammaticality in any strict sense):

- (29) a. \*Geertje isn't coming, isn't she?  
b. \*Geertje isn't coming, is she not?

It is well-known that English polar questions with negation are **biased**, i.e., licit only in contexts with particular configurations of speaker beliefs and/or contextual evidence (Büring and Gunlogson 2000; Sudo 2013, a.m.o.). We propose that these requirements for both high- and low-negation tag Qs conflict with the requirements of uttering a negative declarative; thus, uttering a prosodically-integrated sequence of  $\neg p, \neg p?$  results in systematically defective discourse updates, explaining the badness of (29).

**High-negation tags.** High-negation questions  $\neg p?$  (*Isn't she coming?*) on their own are licensed iff the speaker has a private epistemic bias (approximately, a pre-existing belief) for the positive answer  $p$  (*She is coming*). While this bias is more or less universally agreed to be associated with English HNQs, there is considerable debate about its exact origin (see e.g. Romero

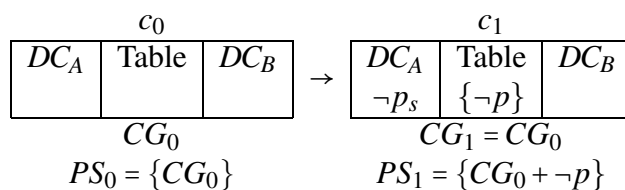
## Polarity Reversal Questions

and Han 2004; Sudo 2013; Frana and Rawlins 2019; Goodhue 2022; Tabatowski 2022). The derivation of such bias is outside the scope of our account, so we make no claims about the specific source of this bias, but assume that wherever it comes from, it is a conventional aspect of the meaning of HNQs.<sup>7</sup>

We take inspiration from a prominent prior analysis of HNQ bias on which a HNQ  $\neg p$ ? expresses the speaker’s certainty that  $p$  should be added to the Common Ground (Romero and Han 2004). We implement that notion in the Table model by making use of an independently-motivated distinction between speaker commitment as *source* vs. *dependent* (Gunlogson, 2008; Malamud and Stephenson, 2015). Sourcehood is a discourse-relative notion: a speaker commits to a proposition  $p$  as source if their commitment introduces evidence for  $p$  into the discourse; if their commitment is based on evidence introduced into the discourse by another source, including someone else’s testimony, that commitment is dependent. Assertions of  $p$  often constitute source-commitments, unless they are reactions to the sudden appearance of evidence for  $p$  in the discourse context; commitments incurred by responding to someone else’s assertion are often (though not necessarily) dependent. We annotate source and dependent commitments with subscript  $s$  and  $d$ , respectively.

We implement the analysis of HNQ bias as expressing the speaker’s certainty that  $p$  should be added to the Common Ground by treating an HNQ  $\neg p$  as presupposing that the speaker is committed to  $p$  as source (viz.,  $p_s \in DC_{Sp}$ ), this being a reasonable translation of that proposal into the Table model.<sup>8</sup> HNQs are otherwise semantically equivalent to corresponding positive polar questions. To see how this will deliver the infelicity of high-negation tags with negative-polarity anchors, consider the case of an update sequence of a negative declarative followed by an HNQ.

- (30) a. He didn’t eat the cake. #Didn’t he eat the cake?  
 b. Utterance of *He didn’t eat the cake* (delivers input context for *Didn’t he eat the cake?*)



The speaker commits themselves to  $\neg p$  in virtue of asserting the declarative  $\neg p$ . We this assume commitment is as source by default ( $\neg p_s$ ), in the absence of a clear target for dependent commitment. (See below for discussion of contexts in which dependent commitment is licensed.) The output context of this update,  $c_1$ , is then the input context for the HNQ update. But the presupposition of the HNQ is not satisfied at  $c_1$ : the speaker is not committed to  $p$  as source. And the presupposition cannot be accommodated without contradiction: to add  $p_s$  to  $DC_A$  via presupposition accommodation would result in the speaker being committed as source to both  $p$  and  $\neg p$ , leveraging nakedly contradictory demands on their SINCERITY. So an HNQ update

<sup>7</sup>Note that this bias is shared between “inner” and “outer” interpretations of HNQs (Ladd, 1981), so potential differences between them are moot; see Sudo (2013).

<sup>8</sup>While we believe it is theoretically desirable to derive the bias of biased questions from their form, namely from their ‘high’ contracted negation, we simply stipulate this as a presupposition for the purposes of this paper; see citations above for various compositional proposals.

is simply not defined relative to a context updated by a corresponding negative declarative.

Because the composed single-step update instantiated by a prosodically incorporated tag question is determined by way of calculating the update sequence carried out by the anchor and tag in sequence, the single-step update carried out by the tag question version of (30), *He didn't eat the cake, didn't he?*, is likewise undefined. Because the presupposition of HNQs is fundamentally incompatible with having asserted  $\neg p$ , we get the systematic unacceptability of negative anchor + HNQ tag questions.

**Low negation tags.** Low negation questions, by contrast with HNQs, do not require speaker epistemic bias for  $p$ . So why should low negation tags be incompatible with negative anchors? Observe that the canonical use case of an LNQ is to reconcile the conflict between evidence for  $\neg p$  and a prior belief that  $p$  (e.g. Buring and Gunlogson 2000; Romero and Han 2004; Gyuris 2017). For instance, (31) is naturally uttered by a speaker whose prior belief is that it is snowing. But epistemic bias is not a strict precondition for felicitous use of LNQs (Romero and Han, 2004). What *is* crucial to license LNQs is a discourse context which supplies good evidence for  $\neg p$  (Buring and Gunlogson, 2000; Romero and Han, 2004; Goodhue, 2022).<sup>9</sup>

- (31) [To someone who is going outside in shorts and a tank top:]  
Is it not snowing?

Just as we did for the commitment-as-source requirement of HNQs, we will treat the restriction that there be compelling evidence for  $\neg p$  available in the discourse context as a presupposition of an LNQ, while leaving open the question of how precisely this presupposition is derived.<sup>10</sup>

This evidential requirement means that LNQ tags should not be compatible with negative anchors. To see why, recall that uttering the declarative  $\neg p$  commits the speaker to  $\neg p_s$ . A following LNQ tag, in virtue of presupposing  $\neg p$  is evident in the discourse context, will enforce alignment between the speaker's commitments and the context, namely that they converge on  $\neg p$ . But if the context and the speaker's private beliefs converge on  $\neg p$ , uttering an interrogative is always infelicitous, since from the speaker's perspective, any issue on the Table containing  $p$  is nonVIABLE from the get-go. Thus, there is no way to both satisfy the presupposition of the LNQ and project a consistent future common ground given the utterance of the negative anchor.

## 7.2. Positive-positive tags

In contrast to negative-negative questions, positive tag questions are compatible with positive anchors, albeit with a different interpretation than a similar PRQ:

- (32) Ernie baked a cake, did he?

Informally, (32) most naturally evokes a context in which the speaker is expressing surprise at

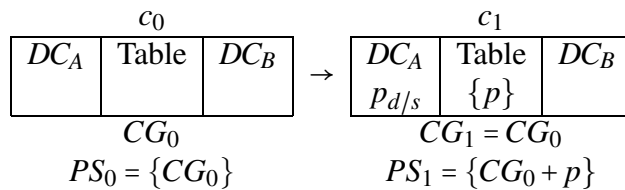
<sup>9</sup>This is a slight oversimplification: LNQs are licit not only in contexts for which there is 'evidence' for  $p$  per se, but also contexts in which the speaker has some particular interest in  $\neg p$  itself, see van Rooy and Šafářová (2003); Romero and Han (2004); Sudo (2013).

<sup>10</sup>Note, anticipating the following section, that this predicts that positive questions will be felicitous only in the absence of compelling contextual evidence for  $\neg p$ , by virtue of MAXIMIZE PRESUPPOSITION!. We take this prediction to be unproblematic.

## Polarity Reversal Questions

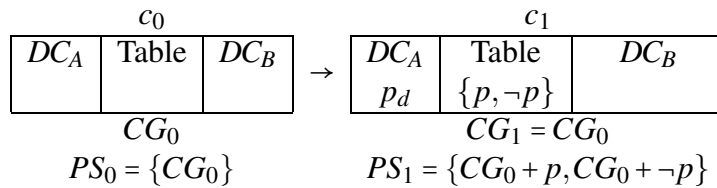
or skepticism of apparent evidence for the proposition *Ernie baked a cake*. We propose that this discourse effect arises via the pragmatic mechanism of antipresupposition inferences (Percus, 2006), deriving from competition between positive and high-negation tags. As analyzed above, an HNQ  $\neg p?$  presupposes that the speaker is committed as source to  $p$ . Given a principle like MAXIMIZE PRESUPPOSITION! (Heim 1991; Lauer 2016), a positive polar question will only be cooperative if the presupposition of its presuppositionally-stronger alternative is false. That is, a positive polar question  $p?$  is felicitous only if the speaker is not committed to  $p$  as source; if they were, they should've chosen the presuppositionally-stronger HNQ. Consider the case of an update sequence of a positive anchor followed by a positive tag:

- (33) a. Ernie baked a cake. Did he eat a cake?  
 b. Utterance of *Ernie baked a cake* (delivers input context for *Did he eat a cake?*)



As mentioned above, assertions often incur commitments as source, but not necessarily. In contexts where evidence for  $p$  has been entered into the discourse by another source—either the assertion of another interlocutor or the observation of a relevant event—an assertion of a declarative sentence can be construed as incurring dependent commitment. The notation  $p_{d/s}$  above represents this ambiguity: prior to the positive polar question follow-up, the utterance of the declarative sentence can incur either dependent commitment or commitment as source, depending on the context. The antipresupposition of a positive polar question resolves this ambiguity, so that in the composed update the speaker's commitment is resolved as being necessarily dependent:

- (34) a. Ernie baked a cake, did he?  
 b. Utterance of *Ernie baked a cake*  $\circ$  utterance of *Did he (bake a cake)?*  
 = Utterance of *Ernie baked a cake, did he?*



By MAXIMIZE PRESUPPOSITION!, the positive polar question *did he?* is felicitous only if the speaker is not committed to  $p$  as source. By the speaker's utterance of the declarative anchor *Ernie baked a cake*, they are committed to  $p$ . If they're committed to  $p$  but not as source, it follows that they have a dependent commitment to  $p$ . The antipresuppositional inference associated with the positive tag forces a dependent interpretation of the commitment incurred by the anchor.

If the speaker's commitment is dependent, there must be something that it is dependent on. That is to say, positive-positive tag questions are only felicitous when the context supplies a source of evidence for  $p$  that the speaker's commitment can be dependent on. This may be an assertion of  $p$  by another party, to which the speaker is responding. Or it might be that the

speaker's utterance is prompted by newly-encountered evidence in the **immediate discourse context**. In such cases, as predicted, positive-positive tags are felicitous, but positive-HNQs are not:

- (35) [Context: Ernie is not known to be a baker. Speaker sees Ernie walking into a potluck with a homemade cake.]  
Ernie baked a cake, did(#n't) he?

The fact that the speaker makes a commitment to  $p$  while still asking whether  $p$  sets up the same pragmatic tension discussed above in the case of PRQs, leading to similar inferences. The fact that the speaker's commitment is dependent in these cases opens up the possibility of an interpretation in which the speaker doesn't fully trust the information relative to which their dependent commitment is licensed, explaining the observation that one prominent use of positive-positive tag questions is the expression of skepticism.

## 8. Conclusion

PRQs wear their discourse effects on their sleeve. They are what they look like they are: an ordinary assertive declarative followed by an ordinary interrogative, prosodically packaged into a single context update. A range of facts about PRQs can be derived from a general treatment of the semantics-pragmatics interface on the level of context update. The variable interpretations of PRQs in different contexts arise from different ways of resolving a pragmatic clash between SINCERITY and VIABILITY. Differences between PRQs and non-integrated sequences result from different inferences generated by performing one composed context update vs two sequential context updates. The polarity restriction can be derived from how that composed context update interacts with the independent bias profile of negative questions.

Finally, though our formalization of prosodic integration is quite general, it is empirically evident that it is subject to some stringent syntactic and pragmatic constraints. For instance, English PRQs seem to require VPE in the tag, and auxiliary identity between the anchor and tag:

- (36) a. \*You're coming, aren't you coming?  
b. \*Belinda is able to waltz, can't she?

These frontiers are yet unexplored, including the cross-linguistic picture. That said, prosodic integration shows promise in explaining the discourse effects of other multi-sentential constructions, such as slifting (Roberts, 2023a).

## References

- AnderBois, S. (2014). On the exceptional status of reportative evidentials. In *Proceedings of SALT 24*, pp. 234–254.
- Asher, N. and B. Reese (2007). Intonation and discourse: Biased questions. In S. Ishihara, S. Jannedy, and A. Schwarz (Eds.), *Interdisciplinary Studies on Information Structure 08*, pp. 1–38. Universitätsverlag Potsdam.
- Biezma, M. and K. Rawlins (2017). Rhetorical questions: Severing asking from questioning. In D. Burgdorf, J. Collard, S. Maspong, and B. Stefánsdóttir (Eds.), *Proceedings of Semantics and Linguistic Theory*, Volume 27, pp. 302–322.



## Polarity Reversal Questions

- Bill, C. and T. Koev (2023). Bias in tag questions. To appear in *Perspectives on Biased Questions*.
- Büring, D. and C. Gunlogson (2000). Aren't Positive and Negative Polar Questions the Same? Ms., UCSC.
- Caponigro, I. and J. Sprouse (2007). Rhetorical questions as questions. In *Proceedings of Sinn und Bedeutung*, Volume 11, pp. 121–133.
- Faller, M. (2019). The discourse commitments of illocutionary reportatives. *Semantics and Pragmatics* 12(8), 1–46.
- Farkas, D. and K. Bruce (2010). On reacting to assertions and polar questions. *Journal of Semantics* 27(1), 81–118.
- Farkas, D. and F. Roelofsen (2017). Division of Labor in the Interpretation of Declaratives and Interrogatives. *Journal of Semantics* 34(2), 237–289.
- Frana, I. and K. Rawlins (2019). Attitudes in discourse: Italian polar questions and the particle *mica*. *Semantics and Pragmatics* 12(16), 1–48.
- Ginzburg, J. (1996). Dynamics and the semantics of dialog. In J. Seligman (Ed.), *Language, Logic, and Computation, Volume 1*, Stanford: CSLI.
- Goodhue, D. (2022). Isn't there more than one way to bias a polar question? *Natural Language Semantics*.
- Groenendijk, J. and M. Stokhof (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph. D. thesis, University of Amsterdam.
- Gunlogson, C. (2001). *True to Form: Rising and Falling Declaratives as Questions in English*. Ph. D. thesis, UC Santa Cruz.
- Gunlogson, C. (2008). A question of commitment. *Belgian Journal of Linguistics* 22, 101–136.
- Gyuris, B. (2017). New Perspectives on Bias in Polar Questions: A Study of Hungarian *-e*. *International Review of Pragmatics* 9(1), 1–50.
- Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language* 10(1), 41–53.
- Han, C.-h. (2002). Interpreting interrogatives as rhetorical questions. *Lingua* 112(3), 201–229.
- Heim, I. (1991). Artikel und definitheit [*Articles and definiteness*]. In A. von Stechow and D. Wunderlich (Eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, pp. 487–535. Berlin: de Gruyter.
- Hirschberg, J. and G. Ward (1995). The interpretation of the high-rise question contour in English. *Journal of Pragmatics* 24(4), 407–412.
- Jeong, S. (2018). Intonation and sentence type conventions: Two types of rising declaratives. *Journal of Semantics* 35(2), 305–356.
- Jun, S.-A. (2022). The ToBI Transcription System: Conventions, Strengths, and Challenges. In J. Barnes and S. Shattuck-Hufnagel (Eds.), *Prosodic Theory and Practice*, pp. 151–211. Cambridge, MA: MIT Press.
- Krifka, M. (2015). Bias in commitment space semantics: Declarative questions, negated questions, and question tags. In S. D'Antonio, M. Moroney, and C. R. Little (Eds.), *Proceedings of SALT 25*, pp. 328–345.
- Ladd, R. D. (1981). A first look at the semantics and pragmatics of negative questions and tag questions. In *CLS 17*, pp. 164–171.
- Lauer, S. (2016). On the status of 'Maximize Presupposition'. In M. Moroney, C.-R. Little, J. Collard, and D. Burgdorf (Eds.), *Proceedings of Semantics and Linguistic Theory 26*, pp.

980–1001.

- Malamud, S. and T. Stephenson (2015). Three ways to avoid commitment: Declarative force modifiers in the conversational scoreboard. *Journal of Semantics* 32(2), 275–311.
- Murray, S. E. and W. Starr (2018). Force and Conversational States. In D. Fogal, D. W. Harris, and M. Moss (Eds.), *New Work on Speech Acts*, pp. 202–236. Oxford University Press.
- Murray, S. E. and W. Starr (2020). The structure of communicative acts. *Linguistics and Philosophy* 44, 425–474.
- Northrup, O. (2014). *Grounds for Commitment*. Ph. D. thesis, UC Santa Cruz.
- Pancheva, R. and D. Rudin (2019). Speaker discourse roles and the discourse profile of reportative evidentials. In *Proceedings of the 22nd Amsterdam Colloquium*, pp. 327–336.
- Percus, O. (2006). Antipresuppositions. In A. Ueyama (Ed.), *Theoretical and empirical studies of reference and anaphora: Toward the establishment of generative grammar as an empirical science*. Japan Society for the promotion of science.
- Roberts, C. (1996/2012). Information structure in discourse. *OSU Working Papers in Linguistics* 49, 91–136.
- Roberts, T. (2023a). There’s no form-meaning mismatch in double negation slifting, i don’t think. In *Ms., University of Amsterdam*.
- Roberts, T. (2023b). Uninquisitive questions. In *Proceedings of IWCS 2023*.
- Romero, M. and C. Han (2004). On negative yes/no questions. *Linguistics and Philosophy* 27(5), 609–658.
- Rudin, D. (2018). *Rising Above Commitment*. Ph. D. thesis, University of California, Santa Cruz.
- Rudin, D. (2022). Intonational commitments. *Journal of Semantics* 39(2), 339–383.
- Sadock, J. (1971). Queclaratives. In *Papers from the 7th regional meeting of the Chicago Linguistics Society*, pp. 223–231.
- Sailor, C. (2009). Tagged for Deletion: A Typological Approach to VP Ellipsis in Tag Questions. Master’s thesis, UCLA.
- Sailor, C. (2014). *The Variables of VP Ellipsis*. Ph. D. thesis, UCLA.
- Scheffler, T. and S. Malamud (2023). “Won’t you?” reverse-polarity question tags in American English as a window into the semantics-pragmatics interface. *Linguistics and Philosophy*.
- Sudo, Y. (2013). Biased polar questions in English and Japanese. In D. Gutzmann and H.-M. Gartner (Eds.), *Beyond Expressives: Explorations in Conventional Non-truth-conditional Meaning*, pp. 277–297. New York: Leiden: Brill.
- Tabatowski, M. (2022). *Preferring to Learn: An Attitudinal Approach to Polar Questions*. Ph. D. thesis, University of Chicago.
- Tottie, G. and S. Hoffmann (2006). Tag questions in British and American English. *Journal of English Linguistics* 34(4), 283–311.
- van Rooy, R. and M. Šafářová (2003). On polar questions. In R. B. Young and Y. Zhou (Eds.), *Proceedings of Semantics and Linguistic Theory 13*, pp. 292–309. Ithaca, NY: CLC Publications.
- Westera, M. (2018). Rising declaratives of the quality-suspending kind. *Glossa: A Journal of General Linguistics* 3, 1–32.