

## Pragmatic representations and online comprehension: Lessons from direct discourse and causal adjuncts

**John Duff**, Department of Linguistics, UC Santa Cruz, Santa Cruz CA, US, [jduff@ucsc.edu](mailto:jduff@ucsc.edu)

**Pranav Anand**, Department of Linguistics, UC Santa Cruz, Santa Cruz CA, US, [panand@ucsc.edu](mailto:panand@ucsc.edu)

**Adrian Brasoveanu**, Department of Linguistics, UC Santa Cruz, Santa Cruz CA, US, [abrsvn@ucsc.edu](mailto:abrsvn@ucsc.edu)

**Amanda Rysling**, Department of Linguistics, UC Santa Cruz, Santa Cruz CA, US, [rysling@ucsc.edu](mailto:rysling@ucsc.edu)

Studies on the reading of appositive relative clauses (ARCs) have found that ARCs seem to exhibit less influence in later parsing and decision-making than similar constructions (Dillon et al., 2014; Dillon et al., 2017), a pattern we call *discounting*. Existing accounts often link discounting to the status of ARCs as independent segments in systems of pragmatic representation. This would predict discounting for other constructions as well. In this study, we test that prediction by investigating the reading of direct discourse speech reports and causal adjuncts in English. Diagnostics supplied by the theoretical literature show that these constructions contribute the same independent segments as ARCs in two different systems of pragmatic organization: direct discourse reports contribute an independent speech act, and causal adjuncts contribute their own discourse units. Nevertheless, in a series of five experiments, we find no evidence of ARC-like discounting for either. We conclude that discounting should not be linked to either of these pragmatic representations, and discuss the outlook for other approaches to the phenomenon.



## 1. Introduction

It is well-known that comprehenders use the representational structure of language as they assign an interpretation to linguistic input in real-time. This has been demonstrated largely for syntactic representations (Dillon et al., 2013; Stowe, 1986; Traxler & Pickering, 1996). But research in formal pragmatics suggests the existence of other, pragmatic representations of sentences: for instance, discourse seems to be organized into segments which address certain issues or goals (Asher & Lascarides, 2003; Grosz & Sidner, 1986; Kehler, 2002; Roberts, 1996/2012). Pragmatic structure has been shown to contribute to online expectations for downstream input (e.g. Clifton & Frazier, 2012), thus conditioning the interpretation of syntactic ambiguities (Rohde et al., 2011), pronouns (Kehler & Rohde, 2013, 2017), and ellipsis (Kroll, 2020).

In this article, we address a line of research which suggests that pragmatic representations can also have a more direct reflex in sentence processing. In a series of experiments comparing appositive relative clauses and restrictive relative clauses, Dillon and colleagues (Dillon et al., 2014; Dillon et al., 2017) have demonstrated that appositive relative clauses undergo a process of online encapsulation we will call *discounting*. Discounted material exerts less influence on downstream parsing and decision-making than other material. These authors suggest that the discounting observed for appositives (when compared to restrictive relatives) is a consequence of their marked pragmatic status. However, as we detail below, the pragmatic contribution of an appositive is quite complex. As a result, it remains unclear what precisely about that contribution leads to discounting.

This paper considers two ways in which appositives are distinct from restrictive relative clauses, each of which could be responsible for discounting: the fact that appositives function as their own *speech act* (Syrett & Koev, 2015), and the fact that appositives function as an independent *discourse unit* in discourse structure, more readily participating in discourse relations (Jasinskaja, 2016; Koev, 2013). We focus on the possibility of discounting separate speech acts or separate discourse units because these hypotheses make the most straightforward predictions: formal pragmatics gives us clear diagnostics for embedded speech acts and discourse units linked by discourse relations, and these hypotheses expect discounting behavior for any construction which meets those diagnostics.

We evaluate each of these two possible routes to discounting in turn by testing for discounting effects with another construction that meets the relevant pragmatic diagnostics. In both cases, we find no evidence of discounting. The first set of studies (Sections 2–5) focuses on the possibility of discounting for separate speech acts. Here, we examine *direct discourse speech reports*, which, like appositive relative clauses, contribute independent speech acts.

(1) Horace said, “**I have no interest in sports.**”

Nevertheless, in three naturalness judgment experiments, we show that direct discourse reports do not behave like appositive relative clauses in terms of discounting. In Sections 6–8, we then turn to the possibility of the discounting of separate discourse units. While there is less theoretical unanimity about what is definitively a separate discourse unit, *causal adjuncts* like the one below are considered separate units by most theories, since they involve an identifiable discourse relation linking them to the matrix clause.

(2) Horace went to bed **because he had no interest in sports**.

In two further naturalness judgment experiments, we show that these causal adjuncts also do not exhibit discounting. We thus conclude that discounting should not be linked to either of these pragmatic representations. At the end of the paper, we reflect on two other possible explanations: another pragmatic approach which makes reference to at-issueness (9.2), and a prosodic approach which makes reference to sentence-sized (implicit) phonological units (9.5).

In the rest of this section, we will begin by considering the existing evidence for discounting, and motivating the need for a more detailed understanding of the phenomenon.

## 1.1 ARCs and pragmatic discounting

English has two distinct classes of relative clauses: *restrictive relative clauses* (RRCs) and *appositive relative clauses* (ARCs).

(3) Restrictive RC: Did the nurses **who were hired in July** receive a housing stipend that month?

(4) Appositive RC: Did the nurses, **who were hired in July**, receive a housing stipend that month?

RRCs provide information which “restricts” the set of entities picked out by the preceding noun phrase, as in the bolded clause in the sentence in (3). In this sentence, the relative clause is what narrows the discussion from all the nurses in the context to just those who were hired in July. In contrast to the restricting function of RRCs, ARCs merely provide additional commentary on the entities introduced by the preceding noun phrases. For instance, in (4), the bolded ARC serves to comment on the entire set of nurses relevant in the context, rather than pick out a subset of them. (This contrast with RRCs is why ARCs are also known as *non-restrictive relative clauses*.)

As examples (3) and (4) show, it is possible for an ARC to contain exactly the same sequence of words as an RRC. In such cases, the only surface-level features which distinguish an ARC are found in prosody (Astruc-Aguilera & Nolan, 2007) or in punctuation, which roughly marks intonational boundaries in the orthographic form.

Crucially, it is also generally accepted that ARCs constitute a separate pragmatic unit from the rest of their sentences (Arnold, 2007; Jasinskaja, 2016; Potts, 2005; Syrett & Koev, 2015). This makes ARCs an ideal testing ground for hypotheses which make predictions about the online influence of pragmatic representations. The particular hypothesis examined in Dillon et al. (2014) and Dillon et al. (2017) is summarized in (5).

- (5) *Pragmatic Discounting*: The online interpretation of natural language depends on the construction of pragmatic representations roughly the size of a sentence. After the complete interpretation of one such unit, the material associated with it is discounted in later parsing and decision-making.

This proposal builds off of the intuitive notion that interpretation proceeds in terms of roughly sentence-sized (i.e. proposition-denoting) units (Grosz & Sidner, 1986). Once a unit of this type is complete and interpreted, representing its internal structure is less important (see, e.g., Potter and Lombardi, 1990, for evidence that structural information is lost in memory after conceptual information is computed). This way of grounding the hypothesis suggests a formalization of discounting in a cue-based retrieval model (e.g. Van Dyke & Lewis, 2003) as activation reduction in memory. Nevertheless, for the purposes of this article, we adopt the term *discounting* to remain neutral as to the particular mechanism.

ARCs and RRCs offer a unique testing ground for this hypothesis. Because ARCs constitute a rare case of a sentence-internal pragmatic segment, Pragmatic Discounting predicts that after an ARC has been read, its content will be represented at a much lower priority than the rest of the sentence, which remains incomplete and relevant to the processing task. In comparison, an RRC in the same sentence does not introduce the same kind of pragmatic segment, and so is not predicted to undergo discounting.

## 1.2 Evidence for the discounting of ARCs

Evidence from online reading measures and offline acceptability judgments seems to support these predictions of Pragmatic Discounting in two ways: (i) ARCs seem to be less influential than RRCs during subsequent dependency resolution in syntactic parsing, and (ii) the complexity of an ARC seems to be less influential on participants' impressions of a sentence's difficulty than the complexity of an RRC

Evidence for the first point comes from Dillon et al. (2017), who investigated online differences in the processing of ARCs and RRCs that intervene in filler-gap dependencies. Participants in their Experiment 1 rated the naturalness of 24 critical items in conditions like those depicted in **Table 1**, which crossed the type of RC in a sentence and the presence or absence of a filler-gap dependency that crossed the RC. Critical gaps in + Filler conditions are indicated as “\_\_”. Dillon and colleagues observed a difference of differences interaction, such that filler-gap dependencies which crossed an RC exhibited a smaller naturalness penalty when that RC was an ARC.

**Table 1:** Sample stimulus from Dillon et al. (2017), Experiment 1.

Structure	-Filler	+ Filler
RRC	The butcher asked if the lady who bought Italian ham was cooking dinner for her guests.	The butcher asked who the lady who bought Italian ham was cooking dinner for __.
ARC	The butcher asked if the lady, who bought Italian ham, was cooking dinner for her guests.	The butcher asked who the lady, who bought Italian ham, was cooking dinner for __.

Eyetracking-while-reading measures for the same items from the authors' Experiment 2 provided evidence that these judgment differences were due to differences in online difficulty – in particular, the cognitive demands of filler retrieval and integration. In go-past times for the final sentence region, Dillon and colleagues observed a slow-down in the +Filler conditions, where the region contained a gap site and thus was the site of filler integration. Crucially, this gap site slow-down was greater for RRC conditions than ARC conditions.

Dillon and colleagues suggested that this systematic variation in reading time penalties was produced by the differential retrieval interference predicted by Pragmatic Discounting: while the content of an RRC serves as a source of interference, ARCs are subject to discounting due to their separate pragmatic organization, and thus contribute less difficulty during retrieval.

In earlier work, Dillon et al. (2014) considered the alternative hypothesis that ARCs are processed in a unique workspace, free from the resource demands generated by the rest of the sentence. An account like this predicts that ARCs themselves would be easier to read than RRCs. However, Dillon et al. (2017) observed no significant differences across multiple measures in the relative clause region itself. This is precisely why they found discounting to be the most credible account, because the critical differences between conditions appeared not in the difficulty of reading RRCs vs. ARCs, but solely in the consequences of having read an RRC vs. an ARC.

Recent investigations into the processing of subject-verb agreement with similar stimuli (Kim & Xiang, 2022; McInnerney & Atkinson, 2020; Ng & Husband, 2017) have provided convergent evidence that ARCs are less influential than RRCs in subsequent dependency resolution. Like filler-gap processing, agreement processing is affected by content that intervenes in the dependency. In particular, plural marking on a verb with a singular subject generally triggers a slowdown, but the size of that slowdown is reduced in the presence of a local plural noun. The effect is known as *agreement attraction*, and is thought to depend on the availability of that local plural *attractor* in memory (Wagers et al., 2009). We might predict that discounting an attractor would reduce the size of an agreement attraction effect. This seems to be borne out: Ng and Husband (2017) and McInnerney and Atkinson (2020) found in self-paced reading studies that attractors within ARCs triggered smaller agreement attraction effects in post-RC verbs than attractors within RRCs. In

contrast, Kim and Xiang (2022), expanding on the design, found that the comprehension of verbs inside ARCs and RRCs was equally affected by pre-RC attractors. This is another instantiation of the asymmetry noted above: preceding content seems to influence the processing of ARCs and RRCs equally, but ARCs seem to have a reduced influence on the processing of subsequent content. As Kim and Xiang (2022) argue, this asymmetry can be explained by a hypothesis where ARCs undergo discounting.

In the first experiment presented here, we build on a related effect in naturalness judgments observed by Dillon et al. (2014). Dillon and colleagues reported a series of judgment experiments on items with RRCs and ARCs of varying complexity.<sup>1</sup> Convergent evidence across multiple experiments suggested that judgments of acceptability showed less sensitivity to the complexity inside a clause when it was an ARC than when it was an RRC. For example, in their Experiment 3, participants rated 24 critical items in the conditions depicted in **Table 2**, which crossed the type of RC present in the item (RRC vs. ARC) and the amount of complexity within it (Long vs. Short). “Long” complexity was created by modifying the subject of the RC with a (sometimes prepositional) object relative clause (ORC).

**Table 2:** Sample stimulus (excerpted) from Dillon et al. (2014), Experiment 3.

Structure	Short	Long
RRC	That evil man who was on the cruise tried to intimidate the waitress.	That evil man who was on the cruise Mary took to the Pacific Islands tried to intimidate the waitress.
ARC	That evil man, the one who was on the cruise, tried to intimidate the waitress.	That evil man, the one who was on the cruise Mary took to the Pacific Islands, tried to intimidate the waitress.

Participants’ ratings displayed another difference of differences interaction, such that the effect of complexity was significantly different across the different RC conditions. Material which decreases naturalness ratings within an RRC was not rated as poorly when it was within an ARC.

As discussed above, these differences do not stem from difficulty reading the RRCs, since the eyetracking data revealed that ARCs and RRCs have the same online profile until after they have been read. Rather, both results can be explained as consequences of discounting in later computation. For Dillon and colleagues’ Experiment 3, the relevant computation is the evidence

---

<sup>1</sup> Note that this experiment uses parenthetical constructions that are better understood as nominal appositives – in this case, always definite descriptions. Our replication, Experiment 1, will do the same, and we will continue to call them ARCs in the main text. We take the difference between nominal appositives and proper ARCs to be unimportant: the pragmatic features of the two constructions are widely assumed to be comparable (Potts, 2005).

integration required to provide a naturalness judgment for a sentence. Participants must ultimately compose their ratings by weighting the naturalness of the stimulus's component parts. It is possible that this weighting process would be sensitive to the same heuristics employed in parsing, in which case material within a discounted unit would exert less influence on judgments than material within the main body of the sentence. With these mediating assumptions in place, we take naturalness rating data on complexity as another source of insight into the same discounting effect observed in online syntactic processing.

These results have been taken as evidence for an account in which material that constitutes a sub-sentential pragmatic unit is heuristically discounted during following computation. By this hypothesis, pragmatic representations have an influence not just in the resolution of meaning but in a variety of lower-level processing computations, including dependency resolution and the construction of acceptability judgments.

Nevertheless, the only evidence that has been marshaled for Pragmatic Discounting has come from the comprehension of ARCs. In the studies reported in this paper, we generate and test predictions about constructions besides ARCs. To do this, we have to consider a series of more specific formulations of Pragmatic Discounting. We will start with an approach inspired by Dillon and colleagues, who have suggested that the discounting of ARCs may derive, in particular, from their status as separate speech acts, minimal units of communicative purpose. To evaluate the predictions of a speech act-based discounting hypothesis, we turn to another construction, direct discourse, which has also been argued to comprise a separate speech act.

## **2. Speech acts and pragmatic structure**

The representational distinctions that have been posited for pragmatic structure do not consist of the same kind of unitary hierarchical structures that are assumed for syntactic representations. Instead, multiple overlapping but distinct ways of organizing input seem to be necessary to capture the variety of computations which a theory of pragmatics must account for. These include, for example, the representation of dependencies between adjacent sentences, and seemingly separable representations which situate language within the human social context. Thus, while Pragmatic Discounting provides a successful account of the data presented in Dillon et al. (2014) and Dillon et al. (2017), to move forward it is necessary to specify which kinds of pragmatic segments undergo discounting.

In this section, we lay the groundwork for such a distinction, which requires a discussion of one type of pragmatic representation, the speech act. We show that the status of ARCs as separate speech acts is shared by another construction, direct discourse, and present evidence that speech act representations are generated during online processing of direct discourse. We ultimately arrive at the prediction that direct discourse will also exhibit discounting, which then motivates the first experiments we present.

## 2.1 A more constrained hypothesis

The key intuition laid out by Dillon et al. (2017) is that a pragmatic unit roughly the size of a sentence serves as the maximal level at which a certain grain of linguistic detail is represented online. We here entertain a particular candidate for that representation.

The resulting hypothesis, stated in (6), is based on the premise that the relevant maximal level at which the processor requires access to internal content is the *speech act*, the minimal linguistic unit which may carry communicative purpose (Austin, 1962; Ross, 1970). Speech acts are assumed to host features required to appropriately model conversation, such as the relationship between an utterance and its author. It is reasonable to assume this is a critical level of meaning-building: in modern commitment-based approaches to discourse in the tradition of Stalnaker (1978), identifying pairs of speech act and speaker within a discourse is necessary to understand the flow of information (Farkas & Bruce, 2010; Farkas & Roelofsen, 2017; Gunlogson, 2001; Murray, 2014). Reference to the speech context is also necessary to resolve the reference of deictic elements like *I*, *you*, and *here* (Kaplan, 1989). Indeed, speaker attribution appears to be accessed quickly for many purposes in online processing, including speaker-specific lexical prediction (Greel et al., 2008), and auditory simulation, to be discussed further in 2.3 (Alexander & Nygaard, 2008; Zhou & Christianson, 2016b; Zhou et al., 2019).

- (6) *Speech Act Discounting*: The online interpretation of natural language depends on the construction of speech act representations in order to compute relationships between the content and the communicative context. After the complete interpretation of a speech act, the material associated with it is discounted in later parsing and decision-making.

Speech acts have rigid enough surface manifestations that they, alongside illocutionary force, have been proposed as a component of syntactic organization itself (Ross, 1970; Speas & Tenny, 2003). If we accept that speech acts correspond to a syntactic unit of a particular size, it would be possible to implement Speech Act Discounting even if comprehenders only generate hierarchical syntactic representations online. Therefore, Speech Act Discounting is a rather parsimonious approach to pragmatic discounting: it assumes a minimal level of representational complexity in our theory of online comprehension.

## 2.2 ARCs are separate speech acts

Speech Act Discounting correctly predicts that ARCs exhibit discounting. As we will review in this section, ARCs instantiate distinct speech acts from their host sentences.

Previous literature (e.g. Syrett & Koev, 2015) bases this conclusion on two key observations: (i) ARCs permit adverbs which are often assumed to be otherwise limited to initial position within an utterance (Potts, 2005; Thorne, 1972), as illustrated in (7a), and (ii) commitment by

the speaker to the proposition that an ARC expresses is unaffected by the illocutionary force of the host utterance (Arnold, 2007; Peterson, 2004), as shown in (7b).<sup>2</sup>

- (7) a. The nurses, **who, confidentially, were hired in July**, received a housing stipend for June.  
 b. Did the nurses, **who were hired in July**, receive a housing stipend?

For both properties, the minimal contrast with RRCs is visible in (8): the sentence in (8a) is ungrammatical, and in the sentence in (8b), the RRC forms part of the terms of the question, rather than providing a parallel comment as in (7b).

- (8) a. \*The nurses **who, confidentially, were hired in July** received a housing stipend for June.  
 b. Did the nurses **who were hired in July** receive a housing stipend?

Syrett and Koev (2015) refer to the property of instantiating a separate speech act as “illocutionary independence.” In addition to independence, some accounts assert that appositives carry unique illocutionary value (AnderBois et al., 2015; Murray, 2014).

We note here that ARCs are pragmatically distinct from RRCs in at least two other ways beyond their status as separate speech acts. First, ARCs are accessible to discourse relations (Jasinskaja, 2016; Koev, 2013), a limited set of asymmetric pragmatic relations that link together two “proposition-sized” chunks of a text (Asher & Lascarides, 2003; Hobbs, 1979, i.m.a.). Those chunks are frequently independent sentences, but ARCs are a conspicuous case of sub-sentential chunks. We’ll call these chunks of meaning which can be linked by discourse relations *discourse units*: ARCs, and not RRCs, therefore can introduce a new discourse unit. We will return to consider this kind of representation in Section 6.

ARCs also provide a different type of information from a typical assertion (AnderBois et al., 2015; Murray, 2014; Potts, 2005; Tonhauser, 2012), which Potts (2005) analyzes as not-at-issue content. Potts suggests that not-at-issue content displays a number of pragmatic hallmarks,

---

<sup>2</sup> We may not be fully convinced by these observations. Fairclough (1973) and Ayres (1974) both contest Thorne’s (1972) claim that ARCs are unique in their ability to host utterance-initial adverbs, showing that RRCs modifying an indefinite seem to be suitable hosts as well (a).

(a) Some nurses who, confidentially, were hired in July received a housing stipend for June.

Observation (ii) is less controversial, but may not require ARCs to be a separate speech act if illocutionary force is encoded as a scope-taking operator like quantification or negation (e.g. Farkas & Roelofsen, 2017; Rizzi, 1997), to which ARCs are also immune. In this approach, we would require only a single theory of projection to account for this behavior, and might not need to label ARC a speech act in itself. (Such an account would, however, still have to resort to speech act-hood if we accept that ARCs have unique illocutionary status of their own, cf. AnderBois et al., 2015; Murray, 2014.) Thus, while we don’t give up the claim of speech act-hood, it should not be taken as innocently as it has been in recent literature.

including inaccessibility to polar response particles like *yes* and *no*, and scopal independence from at-issue content. While at-issueness as a pragmatic feature figures in much of the ARC literature, the extent to which it should be treated as a unified notion is debated (Jasinskaja, 2016; Potts, 2012). We set it aside here, but offer some discussion in 9.2.

Because ARCs have a unique status in all of the levels of pragmatic representation reviewed in this section, the processing evidence in 1.2 is possibly compatible with hypotheses defined over any of them. Nevertheless, we will begin by taking seriously Speech Act Discounting in particular. We move next to discuss another construction involving a sub-sentential speech act, direct discourse speech reports.

### 2.3 Direct discourse and speech acts

Natural language offers us another common pragmatically independent sub-sentential construction in the form of *direct discourse* (DD) speech reports, illustrated in (9). DD should be contrasted with *indirect discourse* (ID), illustrated in (10). These together make up the two primary ways of expressing reported speech as the object of a verb like *say*.

(9) Rebecca said, “**The nurses in my ward are well-paid.**”

(10) Rebecca said **that the nurses in her ward were well-paid.**

(11) a. \*Who did Rebecca say, “**I saw yesterday**”?  
 b. Who did Rebecca say **she saw yesterday**?

Like ARCs and RRCs, these alternatives are minimally different in notable ways. ID encodes only the meaning of reported speech. It is structurally integrated to the extent that it permits binding (*Every nurse<sub>i</sub> said that they<sub>i</sub> were well-paid.*) and sequence of tenses phenomena (*Every nurse said<sub>i</sub> that they were<sub>i</sub> (at that time) well-paid.*) (Cappelen & Lepore, 2017), and in some cases permits Wh-movement (11). In all of this, it resembles a larger class of clause-embedding constructions, cf. *Every nurse thought that she was well-paid.* On the other hand, like ARCs, DD features a non-integrated representation that encodes a non-standard type of information, in this case a mimetic simulation of the reported utterance’s form (Clark & Gerrig, 1990; Davidson, 2015). Notably, in DD that form is entirely opaque to surrounding linguistic material. That is, no structural dependencies of any distance are permitted between DD and its context.

In English, while ID must not be separated from its surrounding context by any punctuation, sentence-medial DD is, by standard prescription, set off by preceding and following commas in addition to raised quotation marks, as illustrated in (12) (e.g. *The Associated Press Stylebook*, 2019; *The Chicago Manual of Style*, 2017).

(12) Pedro said, “We are living in unprecedented times,” and let out a long sigh.

For our purposes, DD is relevant because it, like ARCs, introduces a distinct speech act. The evidence presented above for ARCs as a distinct speech act consisted of two main generalizations: (i) compatibility with a certain set of utterance-modifying adverbs, and (ii) independence from the illocutionary force of an embedding sentence.

Both of these generalizations hold of DD. There is a key difference, though: while an ARC appears to exist as a separate speech act within the ongoing discourse, DD presents a speech act outside of the ongoing discourse. If this distinction matters at all to our theory of speech act organization, it would only make DD a more exemplary candidate as an independent speech act. DD can report anything produced by an event of speech: in addition to permitting utterance-modifying adverbs (13), it can consist solely of non-propositional content, as in (14). Further, no aspect of the embedding material, including illocutionary force, can change any aspect of the speech act presented, as shown in (15).

(13) Rebecca said, “**Confidentially, the nurses in my ward are well-paid.**”

(14) Rebecca said, “**Eek!**”

(15) Did Rebecca say, “**The nurses in my ward are well-paid?**”

ID doesn’t present a speech act in the same way. It does make reference to another speech act’s content, but note that it fails to straightforwardly permit utterance-level modification (16) or non-propositional content (17), and everything it contributes is subject to the illocutionary force of the sentence (18).

(16) #Rebecca said that, **confidentially, the nurses in her ward were well-paid.**

(17) \*Rebecca said that **eek.**

(18) Did Rebecca say that **the nurses in her ward were well-paid?**

DD thus presents a speech act as an excerpt from another conversational context altogether. Beyond this, research on auditory perceptual simulation shows that features of that separate conversational context affect early reading of DD but not ID. The starting point of this work comes from the general finding, independent of DD, that readers appear to mentally simulate qualities of an imagined speaker. In a silent reading experiment reported in Alexander and Nygaard (2008), participants exposed to fast and slow talkers exhibited significantly slower reading times when reading a passage if they were told it was written by the slow talker. Zhou and Christianson (2016a, 2016b) found similar speed modulation, in measures as early as first fixation duration, when participants were asked to imagine a non-native speaker reading a passage. Zhou et al.

(2019) further showed that this simulation resembled the auditory experience of listening to the imagined speaker so much that it led to similar “forgiveness” of grammatical errors (Hanulíková et al., 2012) as indexed by reduced P600s in the ERP record.

This voice simulation has also been observed for participants who read passages of narration which contain direct discourse, but not indirect discourse. In an eye-tracking study, first-pass and go-past reading times of a DD region were modulated by the described speed of the reported speech, while ID showed no effect (Yao & Scheepers, 2011). Even when the speed is described by an immediately pre-verbal temporal adverb indicating the reported speech act was originally delivered quickly, participants have been shown to increase the speed at which they read a passage of DD, and not ID (Stites et al., 2013). fMRI data showed that activity in voice-selective areas of the auditory cortex increased during the reading of DD passages (Yao et al., 2011), indicating that the process modulating reading speed is likely auditory in nature.

In sum, we have strong evidence that the sentence processor is sensitive to distinctions between DD and ID in the early stages of reading, such that DD is treated much like an unembedded piece of text, read as a unit of linguistic material that may have its own distinct speaker and manner of delivery – that is, a speech act. In the features relevant for organizing content into speech acts, then, there is no doubt that DD is distinct from its surrounding context.

In 2.2, we noted that ARCs are argued to be distinguished from RRCs in two other pragmatic ways: their ability to participate in discourse relations, and their contribution of not-at-issue information. It is not clear at first glance that either of these properties distinguish DD and ID in the same way. We will postpone a detailed discussion of the relevant evidence to 6.2.

## 2.4 Predictions

In the subsection above, we reviewed evidence that DD, and not ID, introduces a sub-sentential speech act. Moreover, previous studies have shown this distinction to be active in online comprehension. In this way, DD speech reports are parallel in pragmatic organization to ARCs.

By the Speech Act Discounting hypothesis, discounting patterns are predicted for any separate speech act object, once it is no longer currently being processed. This hypothesis thus predicts that material within a DD report will undergo the same discounting observed for ARCs. Because ID reports are a similar construction which does not introduce a separate speech act, they offer a useful comparison. Given the linking hypotheses discussed in 1.2, Speech Act Discounting thus predicts that complexity within a token of DD will result in a smaller judgment penalty than the same complexity within a token of ID. Likewise, filler-gap dependencies spanning DD will result in a smaller penalty than dependencies spanning ID.

These are exactly the predictions we test in the naturalness judgment experiments reported below. In Experiment 1 (Section 3), we replicate the Complexity  $\times$  Structure interaction between

ARCs and RRCs from Dillon et al. (2014). In Experiment 2 (Section 4), a parallel experiment with DD and ID speech reports, we fail to find the first predicted interaction. In Experiment 3 (Section 5), using the filler-gap design of Dillon et al. (2017), with DD and ID speech reports, we fail to find the second predicted interaction. We take these results as evidence that DD is not discounted in the same manner as ARCs, and thus as evidence against Speech Act Discounting.

### 3. Experiment 1: Replicating appositive discounting

The experiments in Dillon et al. (2014) that first established our effect of interest encompassed 24 items and comparison among 8 conditions, such that each participant only provided 3 observations per condition. In the rest of our experiments here, we tested our predictions about the processing of other constructions in a design using 32 items and comparison among 4 conditions, such that each participant provided 8 observations per condition. In order to maximize our ability to interpret any differences between those results and the established ARC effect, in Experiment 1 we replicated the critical findings of Dillon and colleagues in our expanded design. The results suggest that ARC discounting is a reliable effect, robust to changes in stimuli.

#### 3.1 Method

##### 3.1.1 Materials

Participants read 32 critical items featuring critical sentence-medial subject relative clauses in a  $2 \times 2$  cross of Complexity [Short, Long] and Structure [RRC, ARC]. 24 of the 32 item sets were taken directly from Dillon et al. (2014), and an additional 8 item sets were made according to the same template. In all item sets, just as in Dillon et al. (2014), embedded object relative clauses (ORCs) were used to lengthen the critical relative clauses in the Long conditions. All ORCs featured an inanimate head (e.g. *cruise*) extracted from an object or prepositional object position (e.g. *took \_\_ to the Pacific Islands*), and a proper name subject (e.g. *Mary*). A sample item set is provided in **Table 3**.

**Table 3:** Sample stimulus from Experiment 1.

Structure	Short	Long
RRC	That man who was on the cruise tried to throw a waitress overboard.	That man who was on the cruise Mary took to the Pacific Islands tried to throw a waitress overboard.
ARC	That man, the one on the cruise, tried to throw a waitress overboard.	That man, the one on the cruise Mary took to the Pacific Islands, tried to throw a waitress overboard.

Interleaved with these experimental items, participants saw the 72 fillers used by Dillon et al. (2014) and Dillon et al. (2017), which included a mix of grammatical and ungrammatical cases. The most unnatural items featured verb agreement and reflexive number mismatch, and unlicensed negative polarity items.

### 3.1.2 Participants and procedure

48 native English-speaking participants with at least the equivalent of a high school diploma were recruited on Prolific, and paid in accordance with a \$12/hr wage. Prolific’s user base is global, and we did not restrict participation by nationality. Participants in the studies we report here therefore speak a few different varieties of English. They mostly come from the United States (40%), the United Kingdom (37%), and Canada (10%), but also from South Africa (5%), Australia and New Zealand (3%), Ireland (2%), and other countries (4%).

The experiment itself was scripted using Ibex (Drummond, 2010), and hosted on Drummond’s IbexFarm website, now defunct. After completing a consent form, participants were instructed that they would be rating sentences of English for naturalness on a 7-point scale, using computer keys or mouse. The “1” point was labeled “very unnatural,” and “7” was labeled “very natural.” A few trials of guided practice modeled on the procedure used in Dillon et al. (2014) and Dillon et al. (2017) instructed participants on how to use certain regions of the scale, before they proceeded to further practice. After the final labeled practice item, participants answered questions in a short “burn-in” period consisting of filler items in order to continue familiarizing them with the scale before any critical items were shown.

Critical items were Latin-Squared into 4 lists with equal participation in each, before being randomized and interleaved with the 72 fillers. Mandatory 10 second breaks were interspersed throughout the experiment every 15 trials. At the end of the experiment, participants responded to a short survey about their language experience and their experience on the task before receiving payment.

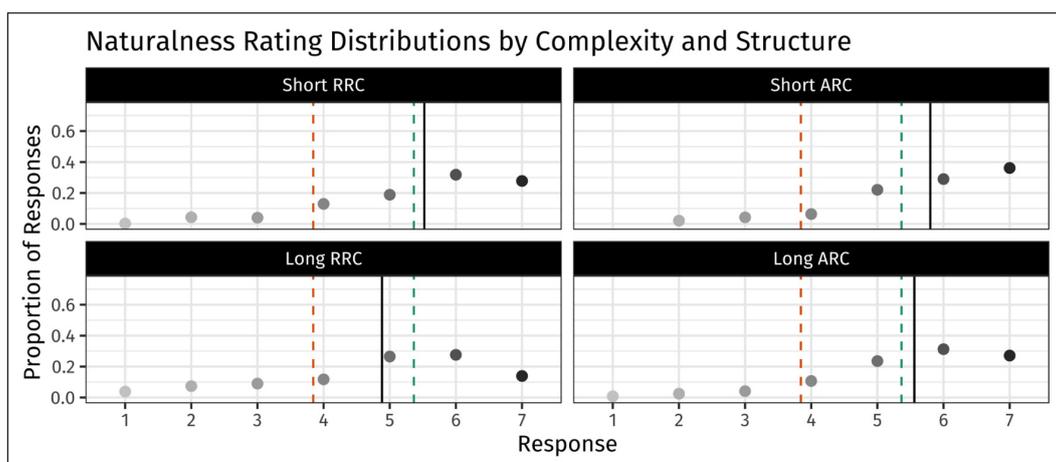
All elements of the procedure for other experiments were identical unless otherwise noted.

## 3.2 Results

Trials featuring latencies longer than 30s were discarded, removing a total of 58 critical trials. The distributions of responses received for each condition are presented in **Figure 1**.

Statistical analysis was carried out by fitting an ordinal mixed-effects model with Stan (Stan Development Team, 2019) using the `brms` package in R (Bürkner, 2017, 2018), with principled weakly-informative priors, maximal random effects, and sum-coded predictors. When we report model results, we present effects with the conditions coded as positive given in parentheses.

Particular model-fitting specifications are summarized in **Table 4**, used throughout the paper unless otherwise specified. Posterior values for  $\hat{\beta}$  and  $\sigma_{\beta}$  along with 95% credible intervals (CRIs) from this model are provided for fixed parameters of interest in **Table 5**. Note that these models treat ordinal data assuming a latent scalar response variable  $(-\infty, \infty)$  with a 0 intercept and a series of fixed thresholds partitioning the underlying space into the available responses (Bürkner, 2020). Fixed effects are slopes for the underlying scalar response variable. We take parameters whose 95% CRI does not contain 0 to indicate noteworthy effects. All models reported feature  $\hat{R} = 1.00$  for the parameters of interest.



**Figure 1:** Naturalness rating distributions by condition from Experiment 1. The black vertical lines report condition means calculated over responses coerced to numerical values, as a rough estimate of central tendency, though we analyze the data as strictly ordinal. Red and green dotted lines report similar means for ungrammatical and grammatical fillers, respectively.

**Table 4:** Model-fitting specifications used in `brms`. All other parameters (e.g. priors for random effects) received their default value.

<b>Family</b>	cumulative ("probit")	<b>Inits</b>	0
<b>Threshold priors</b>	$\mathcal{N}(0,5)$	<b>Chains</b>	6
<b>Fixed slope priors</b>	$\mathcal{N}(0,1)$	<b>Iterations</b>	10000 (incl. 2000 warmup)

We observed a main effect of complexity, such that sentences with long RCs were rated less natural than those with short RCs ( $\hat{\beta} = -0.24$ , 95% CRI =  $(-0.36, -0.13)$ ), and a main effect of structure, such that sentences with ARCs were rated more natural than those with RRCs ( $\hat{\beta} = 0.27$ , 95% CRI =  $(0.11, 0.42)$ ). Most importantly, these main effects were qualified by the

predicted interaction of complexity and structure, such that the complexity penalty was notably smaller for sentences with ARCs ( $\hat{\beta} = 0.08$ , 95% CRI = (0.01, 0.16)).

**Table 5:** Bayesian ordinal mixed-effects model fit to the data from Experiment 1.

Effect	Posterior $\hat{\beta}$	Posterior $\sigma_{\beta}$	95% CRI Lower	95% CRI Upper
Threshold 1 2	-3.19	0.18	-3.54	-2.84
Threshold 2 3	-2.38	0.15	-2.68	-2.08
Threshold 3 4	-1.85	0.15	-2.14	-1.57
Threshold 4 5	-1.20	0.14	-1.49	-0.93
Threshold 5 6	-0.24	0.14	-0.52	0.03
Threshold 6 7	0.93	0.14	0.65	1.21
Structure (ARC)	0.27	0.08	0.11	0.42
Complexity (Long)	-0.24	0.06	-0.36	-0.13
Structure $\times$ Complexity	0.08	0.04	0.01	0.16

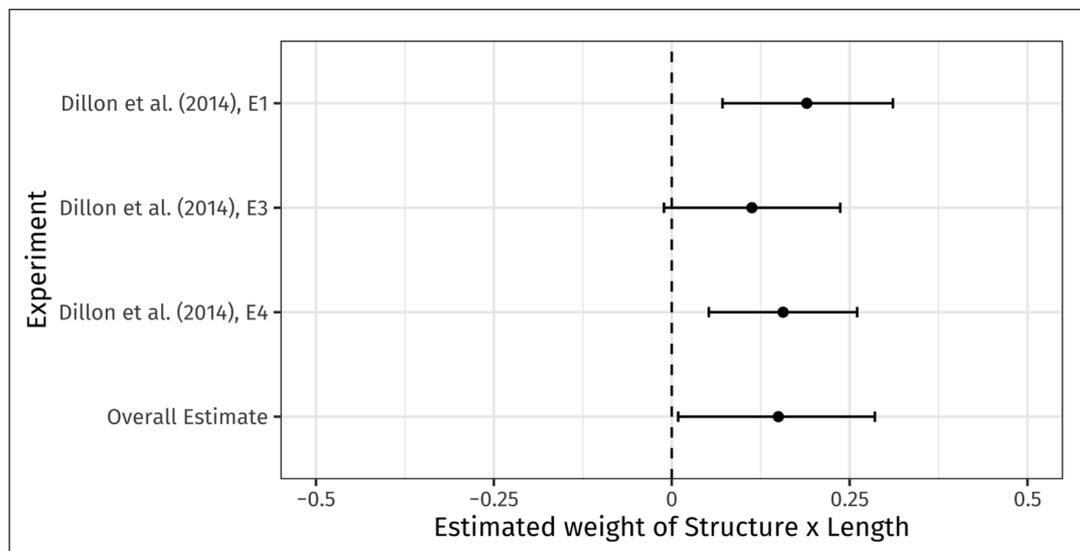
To quantify the degree of evidence in favor of the presence of this critical interaction, we computed a Bayes factor between a fully-specified model and a reduced model without the interaction term,  $BF_{10}$ . Following state-of-the-art recommendations for Bayes factor computation (Nicenboim et al., 2022; Schad et al., 2022), we employed bridge-sampling tools in `brms` (i.e. the function `bayes_factor`), and used empirical priors for the fully-specified model.

To derive these empirical priors, we conducted a meta-analysis of the three studies which have examined naturalness judgment length penalties in RRCs and ARCs in the published literature, Experiments 1, 3, and 4 of Dillon et al. (2014). Applying the same modeling procedure used for the studies we report here, we derived model fits for each of the three experiments. Then for each parameter (the fixed slopes for the main effects and interaction, as well as the six thresholds), we collected the three estimated values and fit to them a final meta-model with a random effect of experiment in order to determine our best expectation for that parameter. Results are compiled in **Table 6**. The estimated weights for the critical interaction in each experiment and the meta-analysis are shown in **Figure 2**.

Using the estimates from our meta-analysis as priors in the fully-specified model for Bayes factor calculation, we arrive at  $BF_{10} = 5.75$  in favor of the model with the interaction. In the taxonomy of Lee and Wagenmakers (2013), the results from Experiment 1 serve as “moderate evidence” for the presence of the critical interaction. We thus replicated the crucial findings of Dillon et al. (2014).

**Table 6:** Parameter expectations induced from meta-analyses over the effects from the Dillon et al. (2014) judgment studies. Fully-specified models used in Bayes factor calculations used these expectations as empirical priors, of the form  $\mathcal{N}(\hat{\beta}, \sigma_{\beta})$ .

Effect	Posterior $\hat{\beta}$	Posterior $\sigma_{\beta}$	95% CRI Lower	95% CRI Upper
Threshold 1 2	-3.15	0.32	-3.79	-2.42
Threshold 2 3	-2.52	0.21	-2.96	-2.08
Threshold 3 4	-1.95	0.19	-2.36	-1.55
Threshold 4 5	-1.36	0.16	-1.67	-1.03
Threshold 5 6	-0.44	0.18	-0.83	-0.06
Threshold 6 7	0.82	0.36	0.05	1.56
Structure (ARC)	0.22	0.26	-0.25	1.05
Complexity (Long)	-0.36	0.08	-0.54	-0.17
Structure $\times$ Complexity	0.15	0.06	0.01	0.29



**Figure 2:** Comparing the estimated weight of the critical Structure  $\times$  Length interaction in the judgment studies reported in Dillon et al. (2014). Error bars display 95% CrIs.

### 3.3 Discussion

In Experiment 1, we replicate the critical effect of Dillon et al. (2014), an interaction in naturalness judgments such that adding complexity to an ARC results in a reduced naturalness penalty, what we call here discounting. We can conclude that the effect is robust, persistent across changes to

the stimuli and sample. Moreover, having replicated the effect with our power, population, and method of analysis in particular, we have established a useful baseline for the results reported in the rest of our study.

## 4 Experiment 2: Costs for complexity within reported speech

In Experiment 2, we conducted an extension of the  $2 \times 2$  from Dillon et al. (2014) discussed and replicated above, which examined naturalness judgments of complexity within DD and ID. The Speech Act Discounting hypothesis predicts that complexity in DD will result in a smaller judgment penalty than ID, due to differences in evidence weighting. This is expected to surface as a difference of differences interaction between Structure and Complexity, such that the penalty associated with Long conditions in ID is attenuated in DD.

### 4.1 Method

We constructed 32 items in a  $2 \times 2$  cross of Complexity [Short, Long] and Structure [ID, DD]. Just as in Dillon et al. (2014), ORCs were used to lengthen the critical regions, here speech report clauses. The particular ORCs were identical to those used in our replication of Dillon and colleagues in Experiment 1. An example item set is given in **Table 7**. DD items included punctuation in accordance with prescriptive standards.

**Table 7:** Sample item set from Experiment 2.

Structure	Short	Long
ID	Evan said that the cruise departed three hours behind schedule.	Evan said that the cruise Mary took to the Pacific Islands departed three hours behind schedule.
DD	Evan said, “The cruise departed three hours behind schedule.”	Evan said, “The cruise Mary took to the Pacific Islands departed three hours behind schedule.”

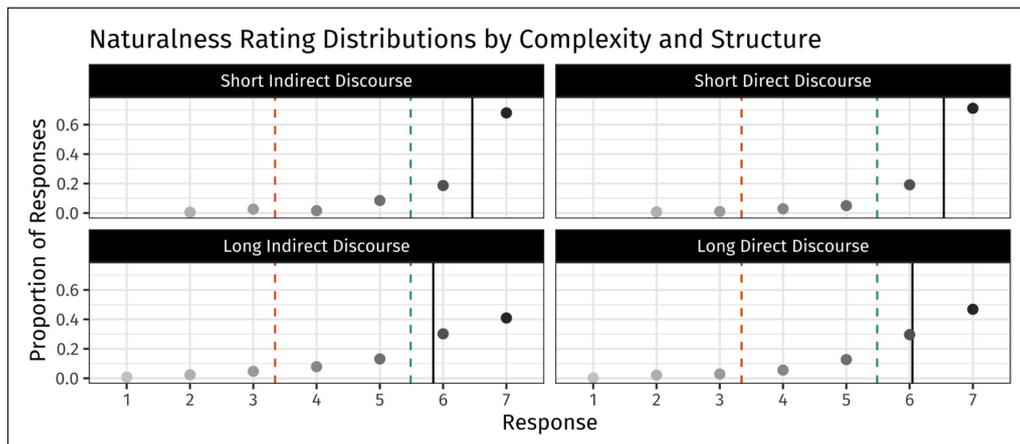
48 participants were recruited and participated as in Experiment 1.

### 4.2 Results

Trials featuring software errors or latencies of longer than 30s were discarded, removing a total of 21 critical trials. The distributions of responses received for each condition are presented in **Figure 3**.

Statistical analysis was performed using a Bayesian-fit ordinal mixed-effects model, reported as above in **Table 8**. We observed a main effect of complexity, such that sentences with long RCs received lower ratings than those with short RCs ( $\beta = -0.51$ , 95% CRI =  $(-0.68, -0.36)$ ). We

report no other noteworthy effects. Most importantly, the interaction of complexity and structure was not credibly non-zero ( $\hat{\beta} = 0.03$ , 95% CRI =  $(-0.05, 0.11)$ ).



**Figure 3:** Naturalness rating distributions by condition from Experiment 2.

**Table 8:** Bayesian ordinal mixed-effects model fit to the data from Experiment 2.

Effect	Posterior $\hat{\beta}$	Posterior $\sigma_{\beta}$	95% CRI Lower	95% CRI Upper
Threshold 1 2	-4.06	0.27	-4.60	-3.55
Threshold 2 3	-3.24	0.21	-3.65	-2.83
Threshold 3 4	-2.70	0.20	-3.09	-2.31
Threshold 4 5	-2.21	0.19	-2.59	-1.83
Threshold 5 6	-1.49	0.19	-1.86	-1.12
Threshold 6 7	-0.31	0.18	-0.67	0.06
Structure (DD)	0.10	0.06	-0.01	0.22
Complexity (Long)	-0.51	0.08	-0.68	-0.36
Structure $\times$ Complexity	0.03	0.04	-0.05	0.11

We computed a Bayes factor using the procedure detailed in Section 3 to quantify the degree of evidence in favor of the absence of the critical interaction ( $BF_{10} = 0.12$ ). This suggests “moderate evidence” in favor of the absence of the critical interaction.

### 4.3 Discussion

In Experiment 2, we failed to find a discounting interaction such that a judgment penalty for complexity within ID is attenuated in DD. The results are inconsistent with Dillon and colleagues’

findings on ARCs, as well as our replication in Experiment 1. We thus fail to support the Speech Act Discounting hypothesis, which would predict DD reports are discounted compared to ID reports just as ARCs are discounted compared to RRCs. This is predicted because DD reports and ARCs both introduce sub-sentential speech acts, and Speech Act Discounting holds that the decision-making process employed in giving naturalness judgments systematically discounts information about sub-sentential speech acts. In contrast to this prediction, DD reports seem to not be discounted at all.

It is apparent that all critical items in Experiment 2 were rated very highly by participants, almost 1 scale degree higher than the comparable relative clause items in Experiment 1. Given that there is a small numeric trend towards the predicted interaction, it is possible that it escaped significance due to scale compression or a ceiling effect. Nevertheless, the results of two further equally-powered follow-ups with slightly different items suggest that these nulls are more than a scale effect. We summarize the findings of these follow-ups briefly below.

In one replication, participants saw the same critical trials and fillers described above, the only difference being 40 additional fillers from another experiment, some of which featured marginally-acceptable split-antecedent NP ellipsis, as in (19).

(19) I saw Mary’s dog and I saw Susan’s cat yesterday, but I didn’t see Jane’s at the time.

We observed no difference in participants’ ratings for the target items, replicating the same main effect of complexity and null interaction as reported above.

In a second replication, we targeted particularly the worry of a scale effect. We adjusted the critical items so that Long conditions featured RCs with animate definite description heads extracted across animate definite description subjects (20), a configuration known to result in exceptional complexity and unacceptability (Traxler et al., 2002).

(20) Evan said, “The captain the senator met at a hotel bar in the Pacific Islands departed three hours behind schedule.”

If the lack of an interaction was due to compression of the main complexity effect, the interaction would have resurfaced when associated with a stronger complexity effect. As predicted, the effect of complexity was larger than in Experiment 1, by a factor of more than two ( $\hat{\beta} = -1.67$ , 95% CRI =  $(-2.00, -1.84)$ ), but estimates for the critical interaction remained close to zero ( $\hat{\beta} = 0.10$ , 95% CRI =  $(-1.34, 0.34)$ ). We are confident that the very low probability of a non-zero interaction parameter is not the result of a scale effect.

## 5. Experiment 3: Dependencies across reported speech

Experiment 3 turned to extending the findings of Dillon et al. (2017), by investigating potential discrepancies in filler-gap processing between ID and DD. This offered another opportunity to

evaluate the predictions of Speech Act Discounting, which predicts that resolving dependencies which cross DD will be less difficult than resolving dependencies which cross ID, by virtue of the processor’s tendency to discount previous speech acts and thereby eliminate sources of interference.

## 5.1 Method

Similar to the construction of materials for Experiment 2, 32 items were created based on the qualities of the 24 items in Dillon et al. (2017). As in that study, each item was manipulated in a  $2 \times 2$  cross of Structure [ID, DD] and Filler [–Filler, + Filler], where the presence of a filler-gap dependency crossing the critical speech report was manipulated by the type of embedded question used, between embedded content questions and embedded polar *if* questions. Speech reports were always contained within relative clauses modifying the subject of the embedded question. An example item is in **Table 9**.

**Table 9:** Sample stimulus from Experiment 3.

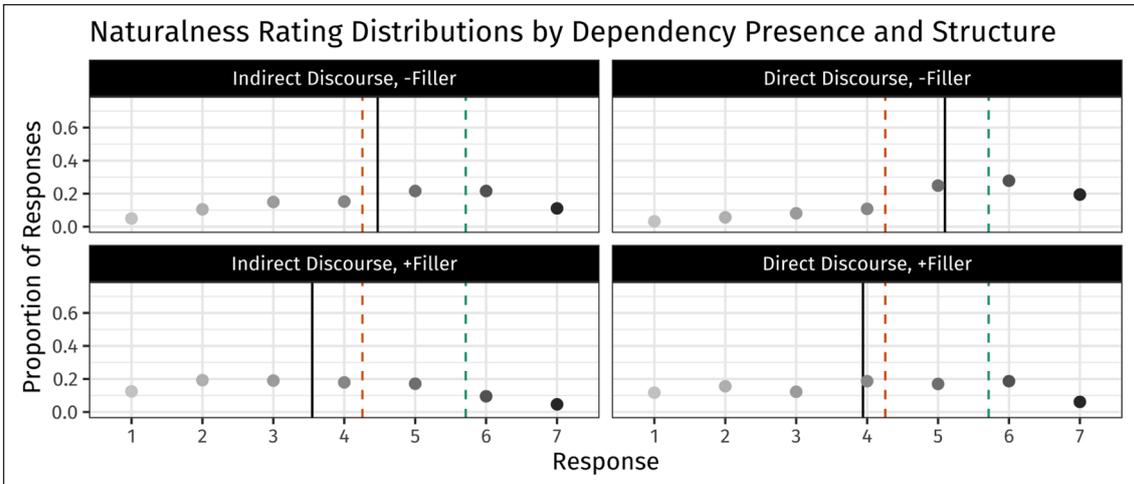
Structure	–Filler	+ Filler
ID	The butcher asked if the lady who said that she would like a nice big ham was cooking for a party.	The butcher asked who the lady who said that she would like a nice big ham was cooking for.
DD	The butcher asked if the lady who said, “I would like a nice big ham,” was cooking for a party.	The butcher asked who the lady who said, “I would like a nice big ham,” was cooking for.

48 participants were recruited and participated as in Experiments 1 and 2.

## 5.2 Results

Trials featuring latencies longer than 30s were discarded, removing a total of 78 critical trials. The distributions of responses received for each condition are presented in **Figure 4**. Statistical analysis was performed using a Bayesian-fit ordinal model, reported as above in **Table 10**.

We observed a main effect (a penalty) of the presence of a filler-gap dependency ( $\hat{\beta} = -0.41$ , 95% CRI =  $(-0.55, -0.26)$ ), and a main effect of structure such that sentences with DD received higher ratings than those with ID ( $\hat{\beta} = 0.19$ , 95% CRI =  $(0.13, 0.26)$ ). The predicted interaction of filler and structure was not credibly non-zero ( $\hat{\beta} = -0.05$ , 95% CRI =  $(-0.12, 0.01)$ ), this time trending numerically in the opposite direction from that predicted by the Speech Act Discounting hypothesis.



**Figure 4:** Naturalness rating distributions by condition from Experiment 3.

**Table 10:** Bayesian ordinal mixed-effects model fit to the data from Experiment 3.

Effect	Posterior $\hat{\beta}$	Posterior $\sigma_{\beta}$	95% CRI Lower	95% CRI Upper
Threshold 1 2	-1.91	0.13	-2.17	-1.65
Threshold 2 3	-1.11	0.12	-1.35	-0.86
Threshold 3 4	-0.54	0.12	-0.78	-0.30
Threshold 4 5	0.02	0.12	-0.21	0.27
Threshold 5 6	0.75	0.12	0.51	1.00
Threshold 6 7	1.74	0.13	1.49	2.00
Structure (DD)	0.19	0.03	0.13	0.26
Filler (+ Filler)	-0.41	0.07	-0.55	-0.26
Structure $\times$ Filler	-0.05	0.03	-0.12	0.01

We computed a Bayes factor using the procedure detailed in Section 3 to quantify the degree of evidence in favor of the absence of the critical interaction. In this case, empirical priors came from a reanalysis of the judgment experiment reported in Dillon et al. (2017), with results reported in **Table 11**. Using these estimates as priors to fit a fully-specified model and comparing it to a reduced model lacking an interaction term yields  $BF_{10} = 0.12$ , suggesting “moderate evidence” in favor of the absence of the critical interaction.

**Table 11:** Bayesian ordinal mixed-effects model fit to the data from Dillon et al. (2017), Experiment 1. The fully-specified model used in the Bayes factor calculations used these expectations as empirical priors, of the form  $\mathcal{N}(\hat{\beta}, \sigma_{\beta})$ .

Effect	Posterior $\hat{\beta}$	Posterior $\sigma_{\beta}$	95% CRI Lower	95% CRI Upper
Threshold 1 2	-2.79	0.20	-3.19	-2.40
Threshold 2 3	-1.94	0.18	-2.29	-1.60
Threshold 3 4	-1.39	0.17	-1.73	-1.06
Threshold 4 5	-0.81	0.16	-1.14	-0.50
Threshold 5 6	-0.005	0.16	-0.32	0.31
Threshold 6 7	1.08	0.17	0.76	1.41
Structure (DD)	0.21	0.08	0.06	0.37
Filler (+ Filler)	-0.54	0.09	-0.72	-0.36
Structure $\times$ Filler	0.06	0.04	-0.03	0.15

### 5.3 Discussion

Just as Experiment 2 failed to find evidence that DD was discounted in judgments, Experiment 3 failed to find evidence that DD presents less retrieval interference. In particular, the corresponding effect with ARCs was revealed in eyetracking to correspond to differential difficulty of filler integration, manifest as faster go-past times on the region containing the gap. The Speech Act Discounting hypothesis predicts that DD, as a speech act distinct from its surrounding context, should also contribute less difficulty – in both cases, the contents of the speech act might be discounted such that they lead to less interference during retrieval of the filler. Under standard assumptions, this would predict a pattern of naturalness judgments parallel to the one reported in Dillon and colleagues’ (2017) Experiment 1. This is not observed: participants’ judgments are impacted by the presence of a report-spanning filler-gap dependency equally for DD and ID reports. In fact, if one were to trust the trends, it would appear that the penalty associated with a filler-gap dependency spanning DD was, contrary to predictions, larger than for one spanning ID. We thus have no evidence that DD is subject to discounting, leaving the predictions of the Speech Act Discounting hypothesis unrealized. The consistency of these null findings lends to their weight.

The fact that we saw no interactions across structures in our experiments is somewhat surprising, given existing experimental work on the reading of DD. As discussed above, readers quickly recognize a passage of DD and modulate their reading rate in accordance with DD’s vividness as a speech report (Alexander & Nygaard, 2008; Stites et al., 2013; Yao & Scheepers,

2011; Zhou & Christianson, 2016b). This would suggest that, from early reading, the processor separates DD as some sort of unit distinct from the ongoing narration, perhaps due to its status as an embedded speech act, and subjects it to special treatment as a result. Nevertheless, this special treatment apparently is not of a type with the unique treatment of ARCs, because it does not include any decrease, or even increase, in the weight afforded to the material within DD in later parsing or decision-making.

## 6. Discourse units and pragmatic structure

Given the lack of advantage for direct discourse over indirect discourse, we do not at present see compelling evidence for Speech Act Discounting. We must then look elsewhere for the source of Dillon and colleagues' results for ARCs. And we would need whatever pragmatic unit discounting is sensitive to to be something that ARCs instantiate, but direct discourse does not. In this section, we consider that this unit may be a *discourse unit*, the atomic information representation within a system of discourse relations like the one outlined by Asher and Lascarides (2003). We show below that available diagnostics suggest that ARCs and not RRCs can contribute sub-sentential discourse units. In contrast, neither DD nor ID speech reports seem able to contribute discourse units in the same way. Unlike a hypothesis at the level of speech acts, which could not distinguish ARCs and DD, a hypothesis at the level of discourse units could explain Dillon and colleagues' results while correctly predicting that DD should not undergo comparable discounting.

In order to test Discourse Unit Discounting, we consider clausal *because* and *when* adjuncts like those below.

(21) Linda was late because she took the train.

(22) Linda was late when she took the train.

The logic of how we arrive at this pair is much more involved than the direct discourse-indirect discourse pair for Speech Act Discounting, so we rehearse it a bit in advance.

Testing Discourse Unit Discounting is a bit more theory-dependent than testing Speech Act Discounting, as the relation between syntactic constructions and discourse units is a source of controversy (Asher, 2000; Buch-Kromann et al., 2011; Dinesh et al., 2005; Hunter, 2016; Jasinskaja, 2016). It is thus non-trivial in practice to find a pair of minimally different constructions where one is definitely a discourse unit and the other definitively is not. In what follows, we will assume that a construction instantiates a separate discourse unit as long as there is a clearly identifiable discourse relation that has been posited to link it to the main clause. In general, as *because* clauses provide explanatory information, they are typically assumed to function as separate discourse units within a sentence, linked by an Explanation relation. The status of temporal adjuncts like (22) is more controversial, but most theories agree that they

are not separate discourse units in at least one case: when they function like a restriction on a quantificational adverbial like *always*, as in the example below.

(23) Charlie always eats well when he's in Europe.

We will thus consider minimal pairs of the following form.

(24) Linda is always early {when, because} she takes the train.

In what follows in this section, we expand on all of these points, first discussing the idea of Discourse Unit Discounting, then considering how direct and indirect discourse fare under that approach, and then turning to the adjunct clauses we experimentally investigate.

## 6.1 Discourse Unit Discounting

To begin, let us assume that the relevant maximal level at which the processor requires access to internal content is the *discourse unit*. Discourse units are typically sentence-size segments of a well-founded discourse (e.g. a narrative, a conversation, etc.) which make a single internally coherent contribution, and might be related to other segments by the discourse relations (coherence relations) of Hobbs (1979) or Asher and Lascarides (2003) (e.g. Narration, Explanation). Reference to such structure is also *a priori* reasonable: much of the existing online evidence for the influence of pragmatic structure on processing comes from examining the online effects of coherence relations between different discourse units – for instance, on anaphora resolution (Kehler and Rohde, 2013, 2017; sources reviewed in Kaiser, 2016) and syntactic parsing (Rohde et al., 2011).

(25) *Discourse Unit Discounting*: The online interpretation of natural language depends on the construction of discourse units in order to compute implicit relations between internally coherent segments of information. After the complete interpretation of a discourse unit, the material associated with it is discounted in later parsing and decision-making.

See also Redeker (2006) for a very similar proposal.

As mentioned briefly in 2.2, ARCs are one of a few constructions which can instantiate sub-sentential discourse units, while RRCs do not do so as easily. We can see this by showing that ARCs alone are accessible for discourse relations. In reviewing this evidence, some terminology is useful: we can speak of discourse relations as relating a *head*, the first discourse unit, and a *tail*, the second discourse unit.

In example (26), from Burton-Roberts (1999, p. 36) (emphasis ours), an ARC, unlike an RRC, appears capable of being the tail of a Result discourse relation with its preceding material. This is associated with the inference that John being employed by the firms frequently is a result of

John getting on with the firms, which is felicitous in example (26b), but does not seem accessible in example (26a).

- (26) a. #John gets on best with those firms **who therefore employ him frequently**.  
 b. John gets on well with those firms, **who therefore employ him frequently**.

An ARC can likewise serve as the head of a discourse relation, with the following sentence being the tail. For example, (27b) shows an ARC heading a Background relation, which supplies the inference that the shopping occurred while and where Lisa was downtown, not while and where she talked to Manuel. RRCs, in contrast, do not yield this interpretation readily, if they do at all: example (27a) does not easily mean that Lisa was shopping for notebooks yesterday.

- (27) a. #Today, Lisa talked to the man **she saw downtown yesterday**. She was shopping for notebooks.  
 b. Today, Lisa talked to Manuel, **who she saw downtown yesterday**. She was shopping for notebooks.

In an account in the framework of Asher and Lascarides (2003), like Jasinskaja (2016), this asymmetry is explained by the fact that ARCs, and not RRCs, introduce a subordinate discourse unit (though cf. Cohen and Kehler, 2021; Hoek et al., 2021; Rohde et al., 2011), and hence are eligible for participation in a discourse relation.

## 6.2 Speech reports and discourse units

One virtue of Discourse Unit Discounting over Speech Act discounting is that it predicts that DD and ID should behave alike, though the reason for the commonality will vary depending on the particular theory of how speech reports are integrated into the discourse structure.

To begin with, we consider approaches where neither DD nor ID instantiate discourse units. Above we showed that subordinate discourse units like ARCs are more accessible for discourse relations with following discourse units than non-discourse units like an RRC. Example (27b) is repeated below as (28). The potential for the final sentence to explain the circumstances of the event in the ARC is derived from the ARC's status as a discourse unit independent from the first part of the sentence.

- (28) Today, Lisa talked to Manuel, **who she saw downtown yesterday**. She was shopping for notebooks.

But DD (29) does not readily permit a parallel interpretation.

- (29) #Today, Lisa said, "**I saw Manuel downtown yesterday**." She was shopping for notebooks.

While ID is somewhat acceptable with the intended interpretation, illustrated in (30), it suggests a context in which the content of the ID is proffered as the main point, the information most

directly answering an implicit Question Under Discussion. For more on this quasi-evidential phenomenon, see Simons (2007); following them, we might assume there is still only a singular discourse unit.

(30) ?Today, Lisa said that **she saw Manuel downtown yesterday**. She was shopping for notebooks.

We might follow this evidence to conclude that, rather than being subordinate discourse units in their own right, DD and ID should be understood merely to furnish part of the semantic content of their matrix-level discourse units (Cumming, 2022). With this in place, Discourse Unit Discounting expects that neither type of speech report would display the discounting phenomena exhibited by ARCs, correctly predicting the failure to find the critical interactions in Experiments 2 and 3.

However, a family of recent pragmatic treatments of speech reports proposes that they do, indeed, involve separate discourse units linked by a discourse relation. Despite this, these too will predict that DD and ID should behave alike. Hunter (2016) and Maier (2020) argue that speech reports feature a particular discourse relation between the fact of the report and the content of the report, Attribution.<sup>3</sup> The evidence for such a proposal mainly comes from cases where the content of speech reports is argued to participate in discourse relations with material outside of the speech report, as in (31). These authors propose that such cases feature three discourse units, corresponding approximately to the propositions (A) *John didn't come to my party*, (B) *Jill said he was out of town*, and (C) *He was out of town*, such that Attribution relates (B) and (C), and Explanation relates (A) and (C). (That is, (C) explains (A).) While Hunter (2016) focuses on ID reports like (31a), Maier (2020) proposes that DD reports like (31b) should receive the same analysis.

- (31) a. John didn't come to my party. Jill said he was out of town. (Hunter, 2016)  
 b. John didn't come to my party. Jill said, "He's out of town."

If DD and ID both involve discourse units, then they should behave the same. Thus, from the perspective of evaluating Discourse Unit Discounting, DD and ID do not form a usefully contrasting pair: If neither DD nor ID instantiate a sub-sentential discourse unit, as we have first considered, we predict that neither will exhibit discounting. If both DD and ID instantiate a sub-sentential discourse unit introduced by an Attribution relation, as proposed by Hunter (2016) and Maier (2020), we predict that both will exhibit discounting. In either case, we correctly predict the absence of an interaction across the two structures in the judgment experiments we have presented in Experiments 2 and 3. While the two different proposals do diverge in terms of

---

<sup>3</sup> We thank an anonymous reviewer for suggesting discussion of these proposals.

their absolute predictions of discounting, it does not seem that we can evaluate these absolute predictions, because discounting phenomena can only be measured relative to a suitable control.

So, while we leave this theoretical question open, we see it as orthogonal to the empirical support for Discourse Unit Discounting. Whether speech reports are simplex or complex in discourse structure, DD and ID share representational properties in a manner consistent with the absence of any discounting effect between them.

### 6.3 Providing positive evidence for Discourse Unit Discounting

Above, we have reviewed evidence that ARCs and RRCs contrast in their status as discourse units, while DD and ID speech reports do not. With these facts in place, Discourse Unit Discounting is consistent with the evidence we've examined. It correctly predicts the interactions in naturalness judgments for ARCs and RRCs observed in Dillon et al. (2014) and Dillon et al. (2017) and replicated in our Experiment 1, and it fails to predict the same interactions for DD and ID speech reports, consistent with our failure to observe interactions in Experiments 2 and 3.

Nevertheless, we should not feel confident in the hypothesis only from negative data: interactions may be absent above for a number of other, theoretically uninteresting reasons. In order to argue for Discourse Unit Discounting in particular, we should establish a piece of positive evidence; that is, we should generate and test a positive prediction.

To do so, we turn to another construction which can contribute a sub-sentential discourse unit, adjunct clauses.<sup>4</sup> While the discourse relations we have discussed so far in this study are all established implicitly, theoretical work commonly suggests that there are various specialized lexical items which explicitly mark certain relations (Asher & Lascarides, 2003; Kehler, 2002; Webber, 2004): e.g. *but* marks Contrast relations, and *because* marks Explanation relations. In examples like (32), the content of the sentence final adjunct *because he's a politician* is thus a discourse unit serving as the tail of an Explanation relation.<sup>5</sup>

(32) George is dishonest because he's a politician.

But not all adjunct clauses pattern in this way. As argued in particular by De la Fuente (2015) (see also De la Fuente et al., 2016), temporal adjunct clauses introduced by *when* exhibit other properties that suggest they do not introduce discourse units of their own. For instance, they presuppose rather than assert the truth of the clause they introduce. Moreover, they are typically understood to describe purely temporal information, situating the time of the events described in the clause with respect to some other events or states. That is, they do not link to the matrix

---

<sup>4</sup> We thank another anonymous reviewer for the suggestion to test our predictions using these constructions.

<sup>5</sup> Whether or not *because* adjuncts contribute a separate speech act is perhaps up for debate. They can indeed host speech act adverbs, but do not exhibit the scopelessness and illocutionary independence discussed for ARCs.

clause via a discourse relation like Narration, Elaboration, or Explanation. However, it has been suggested (e.g. Jasinskaja, 2016) that in some cases they can serve as the tail of a discourse relation like Explanation, leading to the inference of causal sequence in cases like (33), where *when* could be replaced by *because* without notably changing the interpretive link between the clauses.

(33) The squirrel ran away when I shook the broom.

To prevent putative discourse unit interpretations of temporal adjunct clauses, we turn to their use with quantificational adverbials like *usually* and *always*. In such cases, the temporal adjunct clause functions semantically as the restrictor for the quantificational adverbial, providing the domain the quantifier ranges over (see Johnston, 1994; Larson and Sawada, 2012):

(34) George is always dishonest when he's running for office.

For all times  $t$  in which George is running for office at  $t$ , George is dishonest at  $t$

As the rough logical paraphrase in (34) indicates, the clause *when he's running for office* functions to establish the range of times considered for the matrix clause *George is always dishonest*. This is not a function that is readily handled by existing discourse relations. One intuitive reflection of this is the fact that it is difficult to substitute *when* with an expression like *because* without changing the meaning substantially (in particular, breaking the strong semantic connection of the clause as a restrictor of the quantificational adverbial). This quantificational restriction seen in (34) is thus generally taken to be constructed within the semantics, hence preventing *when* clauses from serving as discourse units. Indeed, many theories of discourse, in which discourse relations operate over proposition-denoting discourse units, cannot readily handle the quantification over times in (34).<sup>6</sup>

Given the above discussion, let us take *because* clauses, and not restrictive *when* clauses, to introduce discourse units. We now have a configuration that can test a positive prediction of Discourse Unit Discounting. We predict that just as in comparisons between ARCs and RRCs, *because* clauses will exhibit discounting. In particular, we can test this prediction with the same naturalness judgment design employed by Dillon et al. (2014) and our Experiments 1 and 2. For helpful summary, the predictions of Discourse Unit Discounting across these complexity discounting experiments are summarized in **Table 12**.

---

<sup>6</sup> A notable exception is the Segmented Discourse Representation Theory of Asher and Lascarides (2003), which has the expressive capacity to encode the quantification by *always* at the discourse unit level, though we are unaware of any particular discourse relation that has been advanced to handle cases like (34).

**Table 12:** Predictions of the Discourse Unit Discounting hypothesis.

Target	Control	Prediction	Evidence
ARC	RRC	difference in complexity penalties	✓ (E1)
DD	ID	no difference in complexity penalties	✓ (E2)
<i>because</i>	<i>when</i>	difference in complexity penalties	? (E4)

If *because* clauses are subject to discounting, the judgment penalty observed for added complexity should be reduced compared to restrictive *when* clauses. In Experiments 4 and 5 we tested this prediction, and ultimately find that it is not borne out.

## 7 Experiment 4: Costs for complexity in the subjects of adjunct clauses

### 7.1 Method

As in Experiments 1 and 2, we constructed 32 items in a  $2 \times 2$  cross of Complexity [Short, Long] and Structure [when, because]. A sample item set is given in **Table 13**. The critical adjunct clauses follow present-tense matrix clauses marked with a quantificational adverbial: *often*, *always*, *usually*, or *rarely*. As noted above, in these sentences the *when* clauses will be read as restrictors of these quantifiers, forming part of the truth conditions for the matrix clause. In contrast, the *because* clauses cannot be interpreted as restrictors.

**Table 13:** Sample stimulus from Experiment 4.

Structure	Short	Long
When	Evan often complains to the travel agent when the cruises depart behind schedule.	Evan often complains to the travel agent when the cruises Mary takes to the Pacific Islands depart behind schedule.
Because	Evan often complains to the travel agent because the cruises depart behind schedule.	Evan often complains to the travel agent because the cruises Mary takes to the Pacific Islands depart behind schedule.

In Long conditions, ORCs were again used to add a degree of complexity. While it was generally possible to use the ORCs from Experiments 1 and 2, in 10 item sets lexical changes had to be made to ensure a coherent discourse. A planned item comparison revealed no differences between these 10 item sets and other 22 stimuli, so we do not differentiate between these subsets in our analysis here.

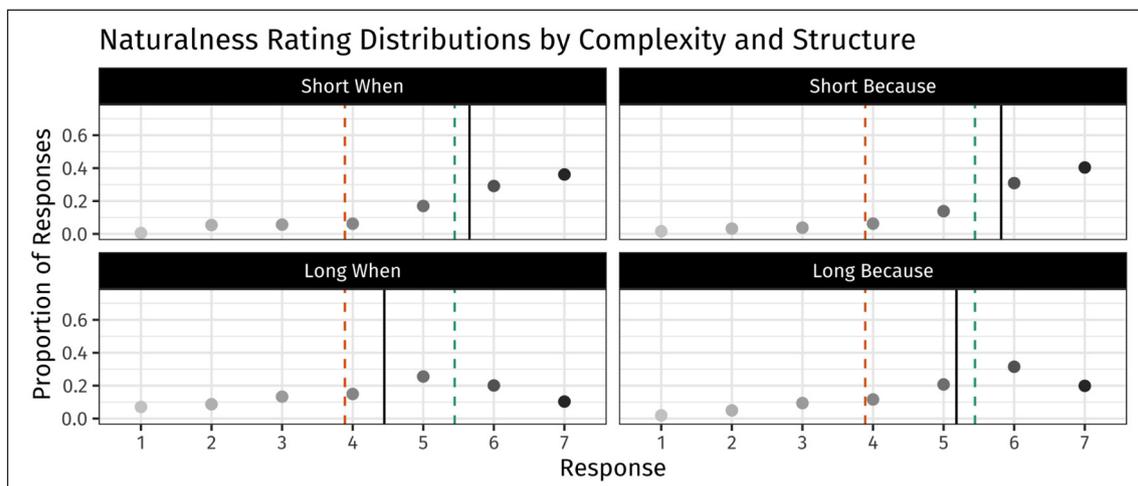
In these items, the *because* clauses permit two possible readings: (i) a low interpretation within the nuclear scope of the quantificational adverbial (e.g. among Evan's complaints, many

are because of late departures), or (ii) a high interpretation outside of the nuclear scope (e.g. the frequency of Evan’s complaints is because of the late departures) (Johnston, 1994; Larson & Sawada, 2012). On the one hand, if we maintain our assumption that *because* always marks a relationship between two discourse units, the scope of quantification will not influence the pragmatic structure. On the other hand, if we give up that assumption, and entertain that material in the nuclear scope of a quantifier is more likely to be part of a single discourse unit, any proportion of these low, single-discourse-unit readings for *because* clauses will make them pattern more with *when* clauses. This crucially could not offer a confounding explanation for the predicted interaction if it is observed, but it could bias against observing the interaction, a question we revisit in the discussion of Experiment 5.

48 participants were recruited and participated as in Experiments 1–3. The experiment was hosted on PClbex (Zehr & Schwarz, 2018), following the closure of IbxFarm. The design, predictions, and analysis plan for this experiment were preregistered.<sup>7</sup>

## 7.2 Results

Trials featuring latencies longer than 30s were discarded, removing a total of 66 critical trials. The distributions of responses received for each condition are presented in **Figure 5**. Statistical analysis was performed using a Bayesian-fit ordinal model, reported as above in **Table 14**.



**Figure 5:** Naturalness rating distributions by condition from Experiment 4.

<sup>7</sup> Available at <https://osf.io/ga3sr>.

**Table 14:** Bayesian ordinal mixed-effects model fit to the data from Experiment 4.

Effect	Posterior $\hat{\beta}$	Posterior $\sigma_{\beta}$	95% CRI Lower	95% CRI Upper
Threshold 1 2	-2.42	0.13	-2.69	-2.16
Threshold 2 3	-1.79	0.12	-2.02	-1.55
Threshold 3 4	-1.29	0.12	-1.52	-1.06
Threshold 4 5	-0.85	0.11	-1.08	-0.63
Threshold 5 6	-0.16	0.11	0.60	1.05
Threshold 6 7	-0.31	0.18	-0.67	0.06
Structure (Because)	0.17	0.04	0.09	0.25
Complexity (Long)	-0.41	0.05	-0.51	-0.31
Structure $\times$ Complexity	0.11	0.03	0.04	0.17

We observed a main effect of complexity, such that sentences with long adjunct clauses were rated less natural than those with short adjunct clauses ( $\hat{\beta} = -0.41$ , 95% CRI =  $(-0.51, -0.31)$ ) and a main effect of structure, such that sentences with *because* clauses were rated more natural than those with *when* clauses ( $\hat{\beta} = 0.17$ , 95% CRI =  $(0.09, 0.25)$ ). Most notably, these main effects were qualified by the predicted interaction of complexity and structure, such that the complexity penalty was notably smaller for sentences with *because* clauses ( $\hat{\beta} = 0.41$ , 95% CRI =  $(0.17, 0.65)$ ). This would seem to be a discounting effect for *because* clauses parallel to that observed in Dillon et al. (2014) and our Experiment 1 for ARCs.

We computed a Bayes factor using the same procedure and priors as in Sections 3 and 4 to quantify the degree of evidence in favor of the critical interaction ( $BF_{10} = 86.74$ ). This suggests “very strong evidence” in favor of the presence of the critical interaction.

### 7.3 Discussion

In Experiment 4, we observe a critical interaction in naturalness judgments such that complexity added to a *because* clause has less influence on participants’ judgments than the same complexity within a *when* clause. This might suggest discounting of *because* clauses in parallel to the discounting of ARCs, and supports the predictions of the Discourse Unit Discounting hypothesis advanced in 6.1. Under that hypothesis, the linguistic material within ARCs and *because* clauses is discounted during judgment processes because they have been parsed into their own discourse units separate from the main content of the sentence. We observe a difference between these former cases and RRCs/restrictive *when* clauses, this account would say, because the latter clauses contribute to the main discourse unit of the sentence, so their content is not discounted, and their complexity thus has an unmitigated effect on judgments.

Nevertheless, there is an alternative explanation for these results that need not say anything about discounting. Following the consensus in formal syntax (Grimshaw, 1977; Hall & Caponigro, 2010), we take *when* clauses to involve relative clause formation within the adjunct clause. In particular, we assume that English *when* clauses are free relative clauses which denote a temporal interval, of the rough meaning ‘at the time that ...’ (see von Stechow & Grønn (2013) for a more formal overview of the semantics). This analysis thus treats *when* clauses alongside more familiar filler-gap constructions (e.g. the argument gaps most familiar in the sentence processing literature), and the centerpiece of that proposal is the observation that *when* clauses show constraints on form (e.g. island effects) and meaning (e.g. structural ambiguity; see Geis, 1970; Larson, 1990) identical to those argued to relate to filler-gap dependencies. For example, free relative sentences like (35b) are thought to be degraded compared to (35a) because they require extraction out of a complex noun phrase (Ross, 1967).

- (35) a. I ate what Mary thought that I should eat \_\_.  
 b. \*I ate what Mary made [the suggestion that I should eat \_\_].

This grammatical contrast is paralleled by an interpretive contrast for *when* clauses:

- (36) a. I ate dinner when Mary (\_\_) thought that I (\_\_) should eat dinner.  
*High Interpretation:* I ate dinner at the time *t* such that Mary thought at *t* that I should eat dinner.  
*Low Interpretation:* I ate dinner at the time *t* such that Mary thought that at *t* I should eat dinner.  
 b. \*I ate dinner when Mary made [the suggestion that I should \_\_ eat dinner].  
*\*Low Interpretation:* I ate dinner at the time *t* such that Mary had previously made the suggestion that at *t* I should eat dinner.

(Hall & Caponigro, 2010, p. 552)

Note first that (36a) is ambiguous between an interpretation where the time of the eating coincides with the time at which Mary is thinking (the High Interpretation) and one where it coincides with the time at which the speaker should eat, in Mary’s opinion (the Low Interpretation). This ambiguity can be explained in structural terms, where the High Interpretation involves a gap in the higher clause and the Low Interpretation one in the lower clause. But then such an explanation predicts that the Low Interpretation should be blocked by island boundaries. And, indeed, example (36b) seems to only have the High Interpretation: as compared to (36a), there is no interpretation available for (36b) where the time that I ate dinner matches the time suggested by Mary.

Crucially, while there is evidence for filler-gap dependencies inside *when* clauses, no such evidence exists for *because* clauses. This could, independent of discourse structural considerations, lead to differences in rating. It is a well-established fact that complexity which intervenes between

a filler and a gap makes the resolution of that dependency particularly difficult (see sources reviewed in Lewis & Vasishth, 2005). For *when* in particular, we expect a filler-gap dependency between its clause-initial position and a gap position adjacent to the verb: Stepanov and Stateva (2015) report self-paced reading evidence for filler-maintenance effects for that span in the processing of *kdaj* ‘when’ in Slovenian. In our experiment, the complexity within *when* clauses may have a greater effect just because it was inserted before the verb and thus interferes with this filler-gap resolution.

In Experiment 5, we evaluate this alternative hypothesis by examining the effect of complexity after the verb, which should not intervene in the dependency instantiated by *when*. We find no evidence for differential complexity costs across adjunct clause types. These results suggest that this effect is indeed explained by filler-gap processing, and does not suffice to provide evidence for Discourse Unit Discounting.

## 8 Experiment 5: Costs for complexity in the objects of adjunct clauses

### 8.1 Method

Using the same design as Experiment 4, we constructed 32 items in a  $2 \times 2$  cross of Complexity [Short, Long] and Structure [*when*, *because*]. Crucially, items were minimally reworked so that the lengthening RC in Long conditions modified the object of the critical adjunct clause. These lengthening RCs were exactly the same as those used in Experiment 4. A sample item set is given in Table 15.

**Table 15:** Sample stimulus from Experiment 5.

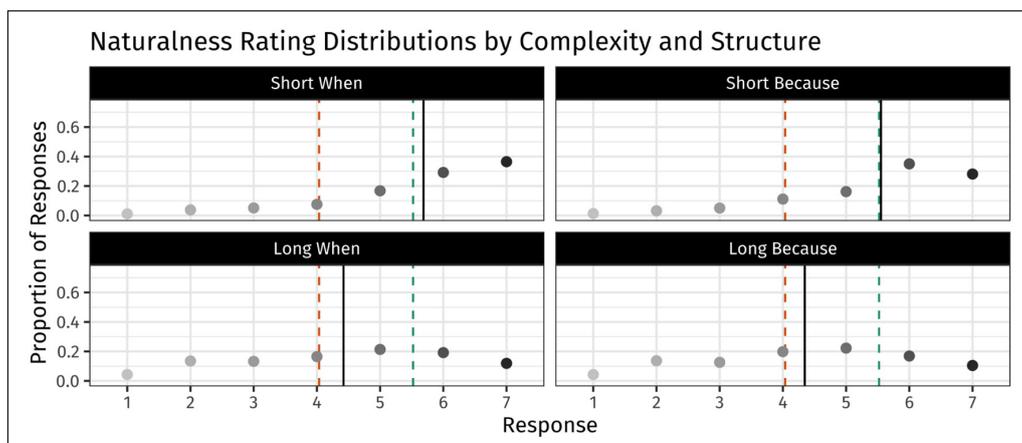
Structure	Short	Long
When	Evan often complains to the travel agent when storms delay the cruises.	Evan often complains to the travel agent when storms delay the cruises Mary takes to the Pacific Islands.
Because	Evan often complains to the travel agent because storms delay the cruises.	Evan often complains to the travel agent because storms delay the cruises Mary takes to the Pacific Islands.

Recall that we assume the filler-gap dependency instantiated by *when* is resolved at the verb. Unlike the materials in Experiment 4, these stimuli do not add complexity until after the verb, so any particular cost observed for the complexity in *when* clauses as compared to *because* clauses could be more confidently attributed to discounting, rather than interference in filler-gap processing.

48 participants were recruited and participated as in the experiments above. Like Experiment 4, the experiment was hosted on PCIBex, and design, predictions, and analysis plan were preregistered.<sup>8</sup>

## 8.2 Results

Trials featuring latencies longer than 30s were discarded, removing a total of 46 critical trials. The distributions of responses received for each condition are presented in **Figure 6**. Statistical analysis was performed using a Bayesian-fit ordinal model, reported as above in **Table 16**.



**Figure 6:** Naturalness rating distributions by condition from Experiment 5.

**Table 16:** Bayesian ordinal mixed-effects model fit to the data from Experiment 5.

Effect	Posterior $\hat{\beta}$	Posterior $\sigma_{\beta}$	95% CRI Lower	95% CRI Upper
Threshold 1 2	-2.67	0.15	-2.97	-2.38
Threshold 2 3	-1.72	0.14	-1.98	-1.45
Threshold 3 4	-1.18	0.13	-1.44	-0.92
Threshold 4 5	-0.58	0.13	-0.83	-0.33
Threshold 5 6	0.13	0.13	-0.12	0.39
Threshold 6 7	1.12	0.13	0.87	1.38
Structure (Because)	-0.06	0.05	-0.16	0.03
Complexity (Long)	-0.53	0.05	-0.63	-0.43
Structure $\times$ Complexity	0.02	0.04	-0.05	0.10

<sup>8</sup> Available at <https://osf.io/hvp2y>.

We observed a main effect of complexity, such that sentences with long adjunct clauses were rated less natural than those with short adjunct clauses ( $\hat{\beta} = -0.53$ , 95% CRI =  $(-0.63, -0.43)$ ). Most strikingly, we failed to observe the predicted interaction of complexity and structure that we observed in Experiment 4 ( $\hat{\beta} = 0.02$ , 95% CRI =  $(-0.05, 0.10)$ ). The Bayes factor, calculated as for the previous experiment, suggests “moderate evidence” in favor of the absence of the discounting interaction ( $BF_{10} = 0.17$ ).

### 8.3 Discussion

In our Experiment 5, we fail to find the critical interaction observed in Experiment 4. In Experiment 4, complexity within the subject of a *because* clause had less of an effect on naturalness judgments than complexity within the subject of a *when* clause. This might have been compatible with a discounting explanation, but we also might expect complexity in the subject of a *when* clause to prompt special difficulty because it intervenes in filler-gap processing. By examining judgment penalties for complexity that could not have intervened in a filler-gap dependency, Experiment 5 offers a way to measure potential discounting without this confound. The lack of an interaction suggests that the interaction in Experiment 4 was driven solely by filler-gap interference with *when*.<sup>9</sup> We are left with no evidence that *because* clauses feature discounting of the kind seen for ARCs, and thus the predictions of Discourse Unit Discounting are not borne out.

Because these items differed from Experiment 4 in the position of added complexity, one might worry that this experiment does not offer a good test for discounting. It is true that complexity in an object position may be simply less influential overall, as it does not hinder processing of the subject-verb dependency like subject complexity. Nevertheless, we still see a robust effect of complexity here, of comparable size to that of Experiment 4. In fact, we’ve already seen discounting effects for object complexity in the original Dillon et al. (2014) studies and our replication in Experiment 1. One might also wonder whether through some unknown mechanism, sentence-final complexity in particular might escape a discounting effect. This does not seem to be the case: Dillon et al. (2014) also report discounting effects for final complexity within a sentence-final ARC.

Finally, given the absence of an interaction, we might return to a possibility flagged above, that despite our assumptions, *because* clauses can sometimes be interpreted locally, without a discourse unit boundary. This might be particularly likely when *because* clauses are interpreted within the scope of a quantifier in their matrix clause, as is possible in some of these items. This kind of variation could be a plausible reason why discounting effects might not be apparent here.

---

<sup>9</sup> Incidentally, we note that this might be a novel finding: though Stepanov and Stateva (2015) discuss the behavior of the Slovenian equivalent, and of other adjuncts in English, filler-gap processing effects have not to our knowledge been investigated for English *when*.

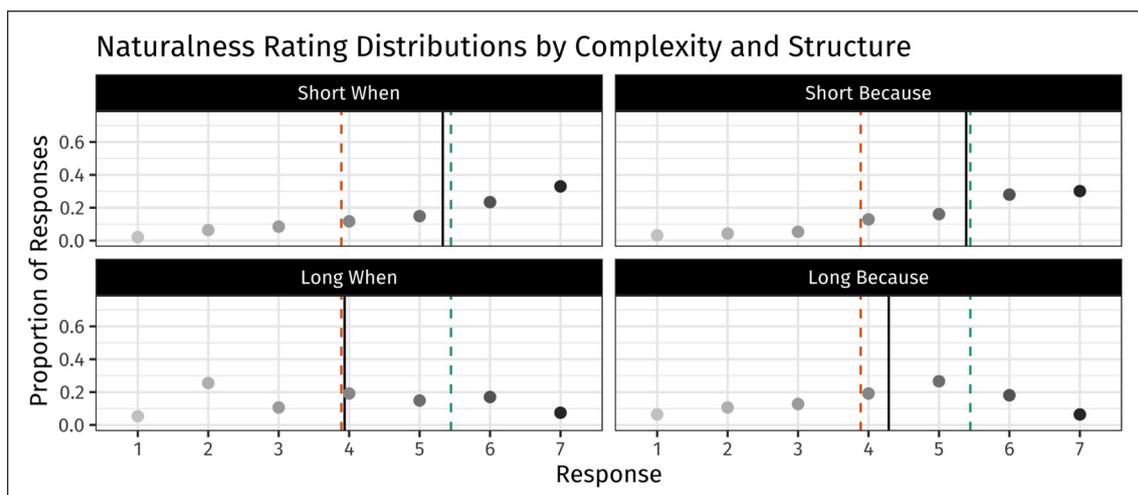
Nevertheless, note first that we expect a general high-scope bias for *because* clauses. Frazier and Clifton (1996) examined scopal ambiguities between *because* clauses and negation, finding that comprehenders experience difficulty when the content of a *because* clause suggests a low-scope interpretation. They conclude that *because* clauses are typically interpreted with high scope, an exception to canonical biases for low-scope interpretation (see also Hemforth & Konieczny, 2004; Koizumi, 2009). From this, we would expect that comprehenders generally opted for high-scope interpretations.

Still, to ensure that our null is not being driven by low-scope interpretations, we examined *post hoc* the ratings given to a subset of our stimuli where low-scope readings should have been impossible. In 8 items that featured the temporal adverb *rarely*, the content of the *because* clause was only ever sensible with a high-scope interpretation, providing the reason for the matrix eventuality's rarity. For instance, in (37), it is implausible that the availability of domestic products is sometimes a reason why Linda purchases European makeup. It is more plausible that this is a reason why Linda generally doesn't buy European makeup.

- (37) Linda rarely purchases European makeup because domestic products are available at the fancy shop (Ralph buys perfume from).

As a result, these items should receive only interpretations where the *because* clause receives high scope. They should therefore be the best chance to observe discounting.

In this *rarely* subset, summarized in **Figure 7**, ratings for items with *because* clauses fell from a mode of 7 to a mode of 5 in the presence of additional complexity. This is roughly the same penalty we observe for *when* clauses throughout.



**Figure 7:** Naturalness rating distributions by condition for the items from Experiment 5 featuring *rarely* in the matrix sentence.

We are left with no evidence that complexity effects are contingent on the interpretation of the *because* clause. We thus take it to be unlikely that the null interaction in this experiment is the result of any variable discourse status of *because* clauses. It would seem instead that causal adjuncts are simply not subject to discounting.

## 9. General discussion

Existing work has demonstrated that certain kinds of material are less impactful than others in parsing and decision-making. The particular constructions that demonstrate this discounting, appositive relative clauses, are notable for the independence of their content from surrounding material. This has inspired the hypothesis that discounting effects are the result of processing mechanisms tied to large pragmatic representations, what we have called here Pragmatic Discounting, repeated below.

- (5) *Pragmatic Discounting*: The online interpretation of natural language depends on the construction of pragmatic representations roughly the size of a sentence. After the complete interpretation of one such unit, the material associated with it is discounted in later parsing and decision-making.

We've attempted to test this hypothesis here by specifying two plausible representations in particular that Pragmatic Discounting might apply over – speech acts and discourse units – and examining whether other sub-sentential constructions that contribute these kinds of representations exhibit discounting. In Experiments 2 and 3, we found no discounting effects for direct discourse speech reports, though they depict independent speech acts, leaving the predictions of our Speech Act Discounting hypothesis unmet. Likewise, in Experiment 5, we found no discounting effects for causal adjuncts, though they instantiate independent discourse units, leaving the predictions of our Discourse Unit Discounting hypothesis unmet. On the whole, these results cast doubt on Pragmatic Discounting as it is framed here. We have ruled out the two most obvious candidates for a pragmatic representation which might undergo discounting.

In the rest of this discussion, we reflect on where this leaves our understanding of pragmatic representations in online processing, and sketch the remaining hypothesis space for an account of discounting.

### 9.1 Online access to pragmatic structure

A pragmatic source for discounting would be of great interest because it would suggest a mechanistic connection between comprehension processes and the representations posited in formal pragmatics. There is a consensus among many psycholinguists that syntactic representations can influence processing in this way, but Pragmatic Discounting would be a

novel case of low-level consequences of speech act or discourse unit organization. We have shown here that existing evidence doesn't merit a conclusion this strong.

We are left with the evidence from other work on pragmatic processing, which is compatible with a more minimal status for pragmatic representations in online processing. Speech acts must be implicated in comprehension to the extent that comprehenders appropriately resolve dependencies between linguistic material and individual speech act participants, as talker simulation evidence suggests they do (Alexander & Nygaard, 2008; Stites et al., 2013; Yao & Scheepers, 2011; Zhou et al., 2019). Likewise, discourse unit parsing must be implicated in comprehension to the extent that comprehenders reference discourse structure to guide pronoun resolution or parsing (Kehler & Rohde, 2013, 2017; Redeker, 2006; Rohde et al., 2011). These kinds of effects are consistent with a world view where participants construct and predict pragmatic structure as one component of a processing mechanism. That component is clearly free to impact other parts of comprehension, but it seems that it doesn't, e.g., provide the representational skeleton for all memory and decision-making processes.

## 9.2 Whither at-issueness?

In Dillon et al. (2014) and Dillon et al. (2017), one of the features of ARCs which is discussed as a potential source for their discounting is their lack of at-issue status. The term comes from Potts (2005), as a label for parts of an utterance's semantic content which are, in that analysis, absent from truth-conditional meaning. Alternative accounts (AnderBois et al., 2015; Murray, 2014; Syrett & Koev, 2015) have, to differing degrees, accepted the label as a primarily pragmatic category of backgrounded information (but cf. Kroll & Rysling, 2019). Often, these accounts aim to capture generalizations which have not been discussed here about the impossibility of denying ARCs with a polar response particle like *no* or using ARCs to answer questions. But other work suggests that the categorical divide between not-at-issue and at-issue meaning may be less important: Jasinskaja (2016), for instance, would reduce this separation to a certain cluster of features associated with the behavior of subordinate discourse units.

One theoretical proposal of not-at-issueness as backgrounding has suggested, in particular, that final ARCs may be "more at-issue", or perhaps more likely to be at-issue, than medial ARCs, as they are diagnosably less backgrounded (Koev, 2013; Syrett & Koev, 2015). On such an account, if not-at-issueness drove discounting, we might expect that final ARCs would be less discounted than medial ARCs. Experiment 3 of Dillon et al. (2014) probes exactly this, and fails to observe any contrast.

Unpublished work from Kroll and Wagers (2017, 2019) followed up on a notion of gradient at-issueness: one might take an ARC which answers a salient question to be more at-issue than one which does not. But Kroll and Wagers found that neither medial vs. final position

nor this kind of salience manipulation had any effect on acceptability judgments. Appositives which were discourse-relevant exhibited discounting just as much as those which were not. This is particularly surprising given that discourse-relevant material is privileged in truth-value judgment tasks (Kroll & Rysling, 2019).

These results suggest that a notion of not-at-issueness based on backgrounding cannot explain the discounting behavior of ARCs. For at-issueness to offer another feasible refinement of Pragmatic Discounting, it would have to be (*contra* Jasinskaja, 2016; Syrett & Koev, 2015) a structural property independent of relevance *per se* (Potts, 2012), one which DD and *because* clauses have and ARCs universally lack. We leave consideration of these premises to future work.

### 9.3 Discounting via visual grouping

ARCs and RRCs have differences outside of the pragmatic domain. Given that we have argued here against a pragmatic source for the discounting of ARCs, we might wonder whether future work could tie discounting to one of these other factors.

A very representationally-conservative approach might return to the visual demarcation of the edges of an ARC via punctuation, in particular, commas. Could the existence of the ARC as an isolated, discounted sub-unit within the sentence be derived directly from this demarcation?

Bottom-up groupings of stimuli by Gestalt principles are recognized early in the visual domain and reflected in memory (Woodman et al., 2003), even when all Gestalt cues in the stimuli are perceived sequentially (Gao et al., 2015). Moreover, commas in particular impact first fixation times during visual sentence processing (Warren et al., 2009), suggestive of their recognition during initial orthographic processing. Nevertheless, the claim that the material bounded by the commas would be discounted would require some *ad hoc* assumptions: typical findings show that grouping assists memory.

Even if such an account could be motivated, it makes the wrong predictions for the patterns observed in our studies. Crucially, a purely grouping-based account of ARC discounting would predict that DD, as another section of text demarcated with punctuation, should be discounted like an ARC. This prediction is not met in Experiments 2 and 3. We suggest, as a result, that a visual grouping approach is not particularly promising.

### 9.4 Discounting via optionality

A reviewer suggests a second account which does not place an explanatory burden on pragmatic structure: it may be possible to derive discounting for ARCs from their syntactic and semantic optionality. While RRCs contribute to the semantic content of their host clauses through restriction, ARCs may be omitted without changing the meaning or grammaticality of their host. A hypothesis built on optionality could also explain the lack of discounting for direct discourse,

compared to indirect discourse: both are arguments of the speech report predicates that embed them, and therefore cannot be omitted.

Nevertheless, we think Experiment 5 provides a strong test against any such hypothesis. *When* and *because* clauses are parallel to RRCs and ARCs in their semantic effects, the former being a form of restrictive modification like an RRC, while the latter may be omitted like an ARC. Despite these structural parallels, the difference we observe in Experiment 1 for the RCs does not obtain in Experiment 5: *because* clauses, though they may be omitted, do not undergo discounting. An approach to discounting relying on optionality could not capture these patterns.

### 9.5 Discounting via prosodic grouping

A final account which may be viewed as “pragmatics-free” could be one based on prosody, the phonological grouping of constituents above the word level. We note that prosody and information structure are often deeply intertwined, and so testing for a prosody-only versus a meaning-only effect may prove extremely difficult, if not impossible. Nonetheless, under a hypothetical purely prosodic explanation, the grouping of lexical constituents in phonological representations could give rise to the differential treatment of certain spans of linguistic material. We think this kind of account should be seriously considered in future research.

Indeed, this kind of approach has been highlighted as a plausible alternative by Dillon et al. (2014) and Dillon et al. (2017). Following evidence for the assignment of implicit prosody to strings in online reading (Fodor, 2002; Frazier et al., 2006), in conjunction with the fact that ARCs tend to instantiate independent units in prosodic representation (Dehé, 2009), we might imagine that discounting holds over large prosodic units rather than pragmatic units. We call this hypothesis Prosodic Discounting (38).

(38) *Prosodic Discounting*: The online interpretation of natural language depends on the construction of (sometimes implicit) phonological representations roughly the size of a sentence. After the complete interpretation of one such unit, the material associated with it is discounted in later parsing and decision-making.

Parallel to our inquiry into Pragmatic Discounting, the validity of Prosodic Discounting rests on several assumptions that will require further investigation. In the first case, we know little about the prosodic parses comprehenders might entertain while reading the six constructions that have been reviewed in this paper. The easiest place to begin would be to understand how speakers of English would explicitly produce such sentences when reading aloud. To relate that data to the online construction of implicit prosodic representations, we need linking hypotheses about how particular phonetic properties correlate with the prosodic representations speakers had in mind. Finally, we need further linking hypotheses about how the representations used in reading aloud correspond to those constructed during silent reading. Because of the complexity of each of these

individual links, reasoning about prosodic intuitions alone will not be sufficient to develop a convincing argument for Prosodic Discounting. If future work is to take this up, we think each of these steps should be considered with care. We linger on a few pieces of relevant research here.

Laboratory studies have indeed established that ARCs demarcated with commas, compared to RRCs without commas, are read aloud with longer prosodic breaks before and after the RC (Astruc-Aguilera & Nolan, 2007), although results are mixed in the absence of punctuation (Hirschberg & Avesani, 1997; Watson & Gibson, 2004). With the appropriate linking hypotheses, these patterns could support the premise that ARCs are comprehended as large prosodic units of their own, and thus subject to Prosodic Discounting.

The present study has added some new desiderata for any account of discounting, beyond ARCs. A feasible Prosodic Discounting account would have to predict a lack of discounting for direct discourse and *because* clauses, compared to indirect discourse and *when* clauses. To determine whether this could be true, we need a finer understanding of the prosody of these constructions. Some corpus data on the production of speech reports finds prosodic breaks at the edges of DD (Jansen et al., 2001), but these are not always clearly present in the signal (Bolden, 2004; Hanote, 2015). Most other work on the production of speech reports focuses on marked excursions from a speaker's typical speech rate and pitch range for both DD and ID (Blackwell et al., 2015; Estelles-Arguedas, 2015; Klewitz & Couper-Kuhlen, 1999; Lampert, 2018); it's not clear whether this should correspond to any difference in prosodic organization. More is known about the prosody of some adjunct clauses. In reading aloud, high-scope *because* clauses generally feature a prosodic break at their left edge (Cooper & Paccia-Cooper, 1980; Hirschberg & Avesani, 2000), in accord with early theoretical intuitions (Downing, 1970).<sup>10</sup> To our knowledge, *when* clauses have not been investigated to the same degree, although most theories predict that *when* should not permit prosodic boundaries at its left edge (e.g. Downing, 1970; Frey & Truckenbrodt, 2015; Selkirk, 2005). If this were true, with some linking assumptions, a Prosodic Discounting hypothesis might incorrectly predict discounting for *because* compared to *when* clauses. We stress that a more careful evaluation is needed to sort these predictions out.

We look forward to future work which can take up these questions and test the predictions of a prosodic account more closely.

## 9.6 Conclusions

We have entertained the hypothesis in this paper that the human processing of language may use pragmatic structure to organize the representation of linguistic material during online computation. We have reviewed the data on appositives from both Dillon et al. (2014) and Dillon

---

<sup>10</sup> Koizumi (2009) suggests that this is also the default implicit prosody for such sentences, as a way of deriving the exceptional high-scope bias for *because*.

et al. (2017) and how it is argued to support this hypothesis. In particular, we have followed up on the proposal that the fine internal content of pragmatic units is discounted after a unit has been completely interpreted.

Here, and elsewhere, a careful understanding of the nature of particular representations can help determine with greater precision the ways in which they might be invoked in processing. Ultimately, in this case, new experiments guided by the theoretical literature suggest that a connection to pragmatic representations may not be appropriate after all. The experiments we have presented here suggest that, counter our predictions, discounting cannot be easily generalized on the basis of the pragmatic structures we have considered.

Nevertheless, we replicate the original discounting effect. The source of the phenomenon, then, remains an open question. We suggest that future work should apply similar scrutiny to other competing explanations, perhaps in particular those based in prosodic representations.

---

## Data accessibility statement

Data, stimuli, analysis scripts, and fitted models are available in a repository hosted by the Open Science Foundation, available at <https://doi.org/10.17605/OSF.IO/5HSXP>.

Designs, predictions, and analysis plans were preregistered for Experiments 4 and 5. Preregistrations can be viewed at <https://doi.org/10.17605/OSF.IO/GA3SR> and <https://doi.org/10.17605/OSF.IO/HVP2Y>.

Experimental materials can be readily tested and copied from versions archived at the PCIBex Farm for Experiment 1 (<https://farm.pcibex.net/r/XlKhdT/>), Experiment 2 (<https://farm.pcibex.net/r/FiWvUS/>), Experiment 3 (<https://farm.pcibex.net/r/UQHhrv/>), Experiment 4 (<https://farm.pcibex.net/r/vNQRyS/>), and Experiment 5 (<https://farm.pcibex.net/r/DCrsXC/>).

## Ethics and consent

All human subjects experimentation reported here was deemed exempt by the UC Santa Cruz Institutional Review Board, under reference number HS1957. All participants provided informed consent.

## Acknowledgements

We thank Lalitha Balachandran, Margaret Kroll, Jess Law, Matt Wagers, Sandy Chung, audiences at CUNY 2020 at UMass Amherst and the Processing Meets Semantics workshop at Utrecht University, and four anonymous reviewers for valuable comments and conversations. Thanks also to Brian Dillon for generously sharing materials and data.

## Competing interests

The authors have no competing interests to declare.

## Author contributions

JD, AR, and PA conceptualized the study and decided on the methodology. AR and PA supervised throughout, and AR acquired funding for participant recruitment. JD conducted the main investigation, and was responsible for data curation and visualization. JD and AB performed the analyses reported here. JD wrote the original draft, and collaborated with AR, PA, and AB to revise and address reviews.

---

## References

- Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: Talker-specific auditory imagery in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 446–459. DOI: <https://doi.org/10.1037/0096-1523.34.2.446>
- AnderBois, S., Brasoveanu, A., & Henderson, R. (2015). At-issue proposals and appositive impositions in discourse. *Journal of Semantics*, *32*, 93–138. DOI: <https://doi.org/10.1093/jos/fft014>
- Arnold, D. (2007). Non-restrictive relatives are not orphans. *Journal of Linguistics*, *43*, 271–309. DOI: <https://doi.org/10.1017/S0022226707004586>
- Asher, N. (2000). Truth conditional discourse semantics for parentheticals. *Journal of Semantics*, *17*, 31–50. DOI: <https://doi.org/10.1093/jos/17.1.31>
- Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.
- Astruc-Aguilera, L., & Nolan, F. (2007). Variation in the intonation of extra-sentential elements. In P. Prieto, J. Mascaró, & M.-J. Sole (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 85–107). John Benjamins. DOI: <https://doi.org/10.1075/cilt.282.08ast>
- Austin, J. L. (1962). *How to do things with words*. Oxford University Press.
- Ayres, G. (1974). I daresay! *Linguistic Inquiry*, *5*(3), 454–456.
- Blackwell, N. L., Perlman, M., & Fox Tree, J. E. (2015). Quotation as a multimodal construction. *Journal of Pragmatics*, *81*, 1–7. DOI: <https://doi.org/10.1016/j.pragma.2015.03.004>
- Bolden, G. (2004). The quote and beyond: Defining boundaries of reported speech in conversational Russian. *Journal of Pragmatics*, *36*, 1071–1118. DOI: <https://doi.org/10.1016/j.pragma.2003.10.015>
- Buch-Kromann, M., Hardt, D., & Korzen, I. (2011). Syntax-centered and semantics-centered views of discourse: Can they be reconciled? In S. Dipper & H. Zinsmeister (Eds.), *Beyond semantics: Corpus-based investigations of pragmatic and discourse phenomena* (pp. 17–30). Bochumer Linguistische Arbeitsberichte.
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. DOI: <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*, 395–411. DOI: <https://doi.org/10.32614/RJ-2018-017>
- Bürkner, P.-C. (2020). *Item response modeling in R with brms and Stan*. arXiv. DOI: <https://doi.org/10.48550/arXiv.1905.09501>
- Burton-Roberts, N. (1999). Language, linear precedence, and parentheticals. In P. Collins & D. Lee (Eds.), *The clause in English: In honor of Rodney Huddleston* (pp. 33–52). John Benjamins. DOI: <https://doi.org/10.1075/slcs.45.05bur>
- Cappelen, H., & Lepore, E. (2017). Quotation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 ed.). Stanford University.

- Clark, H. H., & Gerrig, R. T. (1990). Quotations as demonstrations. *Language*, 66, 764–805. DOI: <https://doi.org/10.2307/414729>
- Clifton, C., Jr, & Frazier, L. (2012). Discourse integration guided by the ‘Question Under Discussion’. *Cognitive Psychology*, 65(2), 352–379. DOI: <https://doi.org/10.1016/j.cogpsych.2012.04.001>
- Cohen, J., & Kehler, A. (2021). Conversational eliciture. *Philosophers’ Imprint*, 21(12). <https://doi.org/2027/spo.3521354.0021.012>
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. De Gruyter. DOI: <https://doi.org/10.4159/harvard.9780674283947>
- Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, 106, 633–664. DOI: <https://doi.org/10.1016/j.cognition.2007.03.013>
- Cumming, S. (2022). Narrative and point of view. In E. Maier & A. Stokke (Eds.), *The language of fiction*. Oxford University Press. DOI: <https://doi.org/10.1093/oso/9780198846376.003.0009>
- Davidson, K. (2015). Quotation, demonstration, and iconicity. *Linguistics and Philosophy*, 38, 477–520. DOI: <https://doi.org/10.1007/s10988-015-9180-1>
- De la Fuente, I. (2015). *Putting pronoun resolution in context: The role of syntax, semantics, and pragmatics in pronoun interpretation* [Doctoral dissertation]. Universite Paris Diderot. DOI: <https://doi.org/10.1515/9783110464108-003>
- De la Fuente, I., Hemforth, B., Colonna, S., & Schimke, S. (2016). The role of syntax, semantics, and pragmatics in pronoun resolution: A cross-linguistic overview. In A. Holler & K. Suckow (Eds.), *Empirical perspectives on anaphora resolution* (pp. 11–31). DOI: <https://doi.org/10.1515/9783110464108-003>
- Dehé, N. (2009). Clausal parentheticals, intonational phrasing, and prosodic theory. *Journal of Linguistics*, 45, 569–615. DOI: <https://doi.org/10.1017/S002222670999003X>
- Dillon, B., Clifton, C., Jr, & Frazier, L. (2014). Pushed aside: Parentheticals, memory and processing. *Language, Cognition and Neuroscience*, 29(4), 483–498. DOI: <https://doi.org/10.1080/01690965.2013.866684>
- Dillon, B., Clifton, C., Jr, Sloggett, S., & Frazier, L. (2017). Appositives and their aftermath: Interference depends on at-issue vs. not-at-issue status. *Journal of Memory and Language*, 96, 93–109. DOI: <https://doi.org/10.1016/j.jml.2017.04.008>
- Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, 69, 85–103. DOI: <https://doi.org/10.1016/j.jml.2013.04.003>
- Dinesh, N., Lee, A., Miltsakaki, E., Prasad, R., Joshi, A., & Webber, B. (2005). Attribution and the (non-)alignment of syntactic and discourse arguments of connectives. In *Proceedings of the Workshop on Frontiers in Corpus Annotation II: Pie in the Sky* (pp. 29–36). DOI: <https://doi.org/10.3115/1608829.1608834>

- Downing, B. (1970). *Syntactic structure and phonological phrasing in English* [Doctoral dissertation]. UT Austin.
- Drummond, A. (2010). *Ibex* [Software]. <https://github.com/addrummond/ibex>
- Estelles-Arguedas, M. (2015). Expressing evidentiality through prosody? Prosodic voicing in reported speech in Spanish colloquial conversations. *Journal of Pragmatics*, 85, 138–154. DOI: <https://doi.org/10.1016/j.pragma.2015.04.012>
- Fairclough, N. (1973). Relative clauses and performative verbs. *Linguistic Inquiry*, 4(4), 526–531.
- Farkas, D. F., & Bruce, K. B. (2010). On reacting to assertions and polar questions. *Journal of Semantics*, 27, 81–118. DOI: <https://doi.org/10.1093/jos/ffp010>
- Farkas, D. F., & Roelofsen, F. (2017). Division of labor in the interpretation of declaratives and interrogatives. *Journal of Semantics*, 34. DOI: <https://doi.org/10.1093/jos/ffw012>
- Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. In B. Bel & I. Marlien (Eds.), *Proceedings of Speech Prosody 2002*.
- Frazier, L., Carlson, K., & Clifton, C., Jr. (2006). Prosodic phrasing is central to language comprehension. *TRENDS in Cognitive Sciences*, 10(6), 244–249. DOI: <https://doi.org/10.1016/j.tics.2006.04.002>
- Frazier, L., & Clifton, C., Jr. (1996). *Construal*. MIT Press.
- Frey, W., & Truckenbrodt, H. (2015). Syntactic and prosodic integration and disintegration in peripheral adverbial clauses and in right dislocation/afterthought. In A. Trotzke & J. Bayer (Eds.), *Syntactic complexity across interfaces* (pp. 75–106). De Gruyter Mouton. DOI: <https://doi.org/10.1515/9781614517900.75>
- Gao, Z., Gao, Q., Tang, N., Shui, R., & Shen, M. (2015). Organization principles in visual working memory: Evidence from sequential stimulus display. *Cognition*, 146, 277–288. DOI: <https://doi.org/10.1016/j.cognition.2015.10.005>
- Geis, M. (1970). *Adverbial subordinate clauses in English* [Doctoral dissertation]. MIT.
- Grimshaw, J. (1977). *English wh-constructions and the theory of grammar* [Doctoral dissertation]. UMass Amherst.
- Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3), 175–2014.
- Gunlogson, C. (2001). *True to form: Rising and falling declaratives as questions in English* [Doctoral dissertation]. UC Santa Cruz.
- Hall, D. P., & Caponigro, I. (2010). On the semantics of temporal *when*-clauses. In *Proceedings of SALT 20* (pp. 544–563). DOI: <https://doi.org/10.3765/salt.v20i0.2566>
- Hanote, S. (2015). Are reporting clauses special cases of parentheticals? In S. Schneider, J. Glikman, & M. Avanzi (Eds.), *Parenthetical verbs* (pp. 257–285). De Gruyter. DOI: <https://doi.org/10.1515/9783110376142-011>

- Hanulikova, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, 24(4), 878–887. DOI: [https://doi.org/10.1162/jocn\\_a\\_00103](https://doi.org/10.1162/jocn_a_00103)
- Hemforth, B., & Konieczny, L. (2004). Scopal ambiguity preferences in German negated clauses. In *Proceedings of CogSci 26* (pp. 559–564).
- Hirschberg, J., & Avesani, C. (1997). The role of prosody in disambiguating potentially ambiguous utterances in English and Italian. In *Proceedings of INT-1997* (pp. 189–192).
- Hirschberg, J., & Avesani, C. (2000). Prosodic disambiguation in English and Italian. In A. Botinis (Ed.), *Intonation: Analysis, modelling and technology*. Springer. DOI: [https://doi.org/10.1007/978-94-011-4317-2\\_4](https://doi.org/10.1007/978-94-011-4317-2_4)
- Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, 3, 67–90. DOI: [https://doi.org/10.1207/s15516709cog0301\\_4](https://doi.org/10.1207/s15516709cog0301_4)
- Hoek, J., Rohde, H., Evers-Vermeul, J., & Sanders, T. J. M. (2021). Scolding the child who threw the scissors: Shaping discourse expectations by restricting referents. *Language, Cognition and Neuroscience*, 36(3), 382–399. DOI: <https://doi.org/10.1080/23273798.2020.1852292>
- Hunter, J. (2016). Reports in discourse. *Dialogue & Discourse*, 7(4), 1–35. DOI: <https://doi.org/10.5087/dad.2016.401>
- Jansen, W., Gregory, M. L., & Brenier, J. M. (2001). Prosodic correlates of directly reported speech: Evidence from conversational speech. In *Prosody in speech recognition and understanding*. International Speech Communication Association.
- Jasinskaja, K. (2016). *Not at issue any more* [Unpublished manuscript]. Department of Philosophy, University of Cologne.
- Johnston, M. J. R. (1994). *The syntax and semantics of verbal adjuncts* [Doctoral dissertation]. UC Santa Cruz.
- Kaiser, E. (2016). Discourse level processing. In P. Knoeferle, P. Pyykkonen-Klauck, & M.W. Crocker (Eds.), *Visually situated language comprehension* (pp. 151–184). John Benjamins. DOI: <https://doi.org/10.1075/aicr.93.06kai>
- Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 565–614). Oxford University Press.
- Kehler, A. (2002). *Coherence, reference and the theory of grammar*. CSLI Publications.
- Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, 39, 1–37. DOI: <https://doi.org/10.1515/tl-2013-0001>
- Kehler, A., & Rohde, H. (2017). Evaluating an expectation-driven Question-Under-Discussion model of discourse interpretation. *Discourse Processes*, 54(3), 219–238. DOI: <https://doi.org/10.1080/0163853X.2016.1169069>
- Kim, S. J., & Xiang, M. (2022). *Memory retrieval selectively targets different discourse units* [Talk]. 35th Annual Conference on Human Sentence Processing, UC Santa Cruz.

- Klewitz, G., & Couper-Kuhlen, E. (1999). Quote-unquote? The role of prosody in the contextualization of reported speech sequences. *Pragmatics*, 9(4), 459–485. DOI: <https://doi.org/10.1075/prag.9.4.03kle>
- Koev, T. K. (2013). *Apposition and the structure of discourse* [Doctoral dissertation]. Rutgers University.
- Koizumi, Y. (2009). *Processing the not-because ambiguity in English: The role of pragmatics and prosody* [Doctoral dissertation]. CUNY.
- Kroll, M. (2020). *Comprehending ellipsis* [Doctoral dissertation]. UC Santa Cruz.
- Kroll, M., & Rysling, A. (2019). The search for truth: Appositives weigh in. In *Proceedings of SALT 29* (pp. 180–200). DOI: <https://doi.org/10.3765/salt.v29i0.4607>
- Kroll, M., & Wagers, M. (2017). *Is working memory sensitive to discourse status? Experimental evidence from responsive appositives* [Poster]. 7th Biennial Experimental Pragmatics Conference (XPrag), University of Cologne. <https://osf.io/95hjq>
- Kroll, M., & Wagers, M. (2019). *Working memory resource allocation is not modulated by clausal discourse status* [Unpublished manuscript]. Department of Linguistics, UC Santa Cruz. [https://people.ucsc.edu/~makroll/uploads/20190201\\_KrollLengthEffectsMs.pdf](https://people.ucsc.edu/~makroll/uploads/20190201_KrollLengthEffectsMs.pdf)
- Lampert, M. (2018). *“Speaking” quotation marks: Toward a multimodal analysis of quoting verbatim in English*. Peter Lang. DOI: <https://doi.org/10.3726/b13406>
- Larson, R. K. (1990). Extraction and multiple selection in PP. *The Linguistic Review*, 7, 169–182. DOI: <https://doi.org/10.1515/tlir.1990.7.2.169>
- Larson, R. K., & Sawada, M. (2012). Root transformations and quantificational structure. In L. Aelbrecht, L. Haegeman, & R. Nye (Eds.), *Main clause phenomena: New horizons* (pp. 47–78). John Benjamins. DOI: <https://doi.org/10.1075/la.190.03lar>
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139087759>
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29, 375–419. DOI: [https://doi.org/10.1207/s15516709cog0000\\_25](https://doi.org/10.1207/s15516709cog0000_25)
- Maier, E. (2020). *Attributions of form and content: A discourse-structural account of reporting* [Unpublished manuscript]. Department of Philosophy, University of Groningen. <https://ling.auf.net/lingbuzz/005597>
- McInnerney, A., & Atkinson, E. (2020). *Syntactically unintegrated parentheticals: Evidence from agreement attraction* [Talk]. 33rd Annual CUNY Conference on Human Sentence Processing, UMass Amherst.
- Murray, S. E. (2014). Varieties of update. *Semantics and Pragmatics*, 7, 2. DOI: <https://doi.org/10.3765/sp.7.2>
- Ng, A., & Husband, E. M. (2017). *Interference effects across the at-issue/not-at-issue divide: Agreement and NPI licensing* [Poster]. 30th Annual CUNY Conference on Human Sentence Processing, MIT.

- Nicenboim, B., Schad, D. J., & Vasishth, S. (2022). *An introduction to Bayesian data analysis for cognitive science*. Advance manuscript [version February 21, 2022]. <https://vasishth.github.io/bayescogsci/book/>
- Peterson, P. (2004). Non-restrictive relatives and other non-syntagmatic relations in a lexical-functional framework. In M. Butt & T. H. King (Eds.), *Proceedings of LFG 2004*.
- Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, 29, 633–654. DOI: [https://doi.org/10.1016/0749-596X\(90\)90042-X](https://doi.org/10.1016/0749-596X(90)90042-X)
- Potts, C. (2005). *The logic of conventional implicatures*. Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199273829.001.0001>
- Potts, C. (2012). Conventional implicatures and expressive content. In C. Maienborn, K. von Stechow, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (pp. 2516–2536). De Gruyter. DOI: <https://doi.org/10.1515/9783110589849-017>
- Redeker, G. (2006). Discourse markers as attentional cues at discourse transitions. In K. Fischer (Ed.), *Approaches to discourse particles: Studies in pragmatics* (pp. 339–358). Elsevier. DOI: [https://doi.org/10.1163/9780080461588\\_019](https://doi.org/10.1163/9780080461588_019)
- Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar*. Kluwer. DOI: [https://doi.org/10.1007/978-94-011-5420-8\\_7](https://doi.org/10.1007/978-94-011-5420-8_7)
- Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics*, 5, 1–69. (Original work published 1996). DOI: <https://doi.org/10.3765/sp.5.6>
- Rohde, H., Levy, R., & Kehler, A. (2011). Anticipating explanations in relative clause processing. *Cognition*, 118(3), 339–358. DOI: <https://doi.org/10.1016/j.cognition.2010.10.016>
- Ross, J. R. (1967). *Constraints on variables in syntax* [Doctoral dissertation]. MIT.
- Ross, J. R. (1970). On declarative sentences. In R. A. Jacobs & P. S. Rosenbaum (Eds.), *Readings in English transformational grammar* (pp. 222–277). Ginn and Company.
- Schad, D. J., Nicenboim, B., Burkner, P.-C., Betancourt, M., & Vasishth, S. (2022). Workflow techniques for the robust use of Bayes factors [Advance online publication]. *Psychological Methods*. DOI: <https://doi.org/10.1037/met0000472>
- Selkirk, E. (2005). Comments on intonational phrasing in English. In S. Frota, M. Vigario, & M. J. Freitas (Eds.), *Prosodies* (pp. 11–58). De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110197587.1.11>
- Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua*, 117, 1034–1056. DOI: <https://doi.org/10.1016/j.lingua.2006.05.006>
- Speas, M., & Tenny, C. (2003). Configurational properties of point of view roles. In A. M. Di Sciullo (Ed.), *Asymmetry of grammar* (pp. 315–344). John Benjamins. DOI: <https://doi.org/10.1075/la.57.15spe>
- Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics 9* (pp. 78–95). Academic Press. DOI: [https://doi.org/10.1163/9789004368873\\_013](https://doi.org/10.1163/9789004368873_013)

- Stan Development Team. (2019). Stan modeling language users guide and reference manual, version 2.26. <https://mc-stan.org>
- Stepanov, A., & Stateva, P. (2015). Cross-linguistic evidence for memory storage costs in filler-gap dependencies with wh-adjuncts. *Frontiers in Psychology*, 6, 1301. DOI: <https://doi.org/10.3389/fpsyg.2015.01301>
- Stites, M. C., Luke, S. G., & Christianson, K. (2013). The psychologist said quickly, “Dialogue descriptions modulate reading speed!” *Memory and Cognition*, 41, 137–151. DOI: <https://doi.org/10.3758/s13421-012-0248-7>
- Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1(3), 227–245. DOI: <https://doi.org/10.1080/01690968608407062>
- Syrett, K., & Koev, T. (2015). Experimental evidence for the truth conditional contribution and shifting information status of appositives. *Journal of Semantics*, 32, 525–577. DOI: <https://doi.org/10.1093/jos/ffu007>
- The Associated Press Stylebook*. (2019). Associated Press.
- The Chicago Manual of Style* (17th ed.). (2017). University of Chicago Press.
- Thorne, J. P. (1972). On nonrestrictive relative clauses. *Linguistic Inquiry*, 3(4), 552–556.
- Tonhauser, J. (2012). Diagnosing (not-)at-issue content. In *Proceedings of SULA 6* (pp. 239–254).
- Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, 47, 69–90. DOI: <https://doi.org/10.1006/jmla.2001.2836>
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, 35, 454–475. DOI: <https://doi.org/10.1006/jmla.1996.0025>
- Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalysed ambiguities. *Journal of Memory and Language*, 49, 285–316. DOI: [https://doi.org/10.1016/S0749-596X\(03\)00081-0](https://doi.org/10.1016/S0749-596X(03)00081-0)
- von Stechow, A., & Gronn, A. (2013). Tense in adjuncts part 2: Temporal adverbial clauses. *Language and Linguistics Compass*, 7(5), 311–327. DOI: <https://doi.org/10.1111/lnc3.12019>
- Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61, 206–237. DOI: <https://doi.org/10.1016/j.jml.2009.04.002>
- Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z Reader. *Cognition*, 111(1), 132–137. DOI: <https://doi.org/10.1016/j.cognition.2008.12.011>
- Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19(6), 713–755. DOI: <https://doi.org/10.1080/01690960444000070>

- Webber, B. (2004). D-LTAG: Extending lexicalized TAG to discourse. *Cognitive Science*, 28, 751–779. DOI: <https://doi.org/10.1016/j.cogsci.2004.04.002>
- Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, 10(1), 80–87. DOI: <https://doi.org/10.3758/BF03196470>
- Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience*, 23(10), 3146–3152. DOI: [https://doi.org/10.1162/jocn\\_a\\_00022](https://doi.org/10.1162/jocn_a_00022)
- Yao, B., & Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition*, 121, 447–453. DOI: <https://doi.org/10.1016/j.cognition.2011.08.007>
- Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)* [Software]. DOI: <https://doi.org/10.17605/OSF.IO/MD832>
- Zhou, P., & Christianson, K. (2016a). Auditory perceptual simulation: Simulating speech rates or accents? *Acta Psychologica*, 168, 85–90. DOI: <https://doi.org/10.1016/j.actpsy.2016.04.005>
- Zhou, P., & Christianson, K. (2016b). I “hear” what you’re “saying”: Auditory perceptual simulation, reading speed, and reading comprehension. *Quarterly Journal of Experimental Psychology*, 69(5), 972–995. DOI: <https://doi.org/10.1080/17470218.2015.1018282>
- Zhou, P., Garnsey, S., & Christianson, K. (2019). Is imagining a voice like listening to it? Evidence from ERPs. *Cognition*, 182, 227–241. DOI: <https://doi.org/10.1016/j.cognition.2018.10.014>

