Proceedings of SALT 35: 583–607, 2025

Incremental interpretation of discourse coherence: Evidence from reading times*

John Duff
UCLA

Kelsey Sasaki
Pomona College

Daniel Altshuler
University of Oxford

Abstract In some discourses, a given inference may be natural at one point, but impossible at the next. We focus here on such cases with inferences of temporal and causal order usually described as ambiguities of discourse coherence. The fact that these inferences come and go over the course of a discourse raises challenges for their representation within theories of dynamic semantics: representations must either include coherence-related meaning in a way that can be selectively edited later, or else persistently underspecify it. As theories of humans' actual cognitive states, these theories make different predictions for patterns of difficulty during real-time comprehension. Borrowing a standard assumption from sentence processing, true incremental representation should be associated with some detectable REGRET when a comprehender meets input incompatible with a previously-preferred analysis. Previous work on the processing of coherence has failed to demonstrate REGRET, suggesting that underspecification is always the strategy at the discourse level, unlike what we find in syntactic processing. In a new study, we probe further, by constructing discourses with a strong, verified bias towards an initial inference, and testing the effects of cues which can reverse this inference early in the discourse, or much later. We observe that these discourses provoke REGRET in self-paced reading, in particular when the critical cue comes late, in a discourse unit following the initial ambiguity. Among many formal accounts which could explain this pattern, we elaborate in particular on a version of Segmented Discourse Representation Theory where analyses are ranked at the offset of each discourse unit.

Keywords: discourse coherence, underspecification, pragmatic processing, causal inference, temporal reasoning, garden paths, SDRT

* We thank Asya Achimova, Oliver Bott, Adrian Brasoveanu, Jesse Harris, Dag Haug, Jet Hoek, Elsi Kaiser, Christina Kim, Alex Lascarides, Dan Lassiter, Hannah Rohde, Julian J. Schlöder, and reviewers and audiences at Amsterdam Colloquium 2024, DiscoMatiX @ XPrag 2025, HSP 38, the "Meaning and grammar" research group at Edinburgh, RED 2025 and SALT 35. This research was supported by a John Fell Fund grant to JD and DA, and a Leverhulme ECF-2022-604 to KS.

1 Introduction

In some sequences of discourse units $[\pi_a \ldots \pi_n, \pi_{n+1}]$, the natural interpretation that would be given for $[\pi_a \ldots \pi_n]$ alone is not part of the natural interpretation of the entire discourse $[\pi_a \ldots \pi_n, \pi_{n+1}]$. For instance, given only π_1–π_3 of (1), π_3 likely describes an event which preceded, and caused, Gary's demotion. But given π_4, the skipping of the meeting must have followed the demotion, as a reaction.

(1) [There has been some chaos recently on the sales team.]$_{\pi_1}$ [Stella demoted Gary.]$_{\pi_2}$ [He skipped a meeting,]$_{\pi_3}$ [because the demotion angered him.]$_{\pi_4}$

This is a consequential observation for dynamic semantic theories that aim to capture how meaning accumulates with each new discourse unit in a sequence.[1] For such theories, the first representation of π_1 through π_3 must be a step along the construction of the final meaning of the discourse. If that step includes a representation of the inferred temporal order of the events in π_2 and π_3, or their inferred causal relationship, that representation must somehow be revisable.

Lascarides & Asher (1993) offer one way to selectively permit these kinds of revisions. They make a representational distinction between the core semantic content of a sequence, which is updated monotonically (i.e., without revision), and discourse-pragmatic meaning, which may be freely (re-)computed or (re-)ranked at each step given defeasible heuristics. In this system, while the initial inferences of causality and temporal order are part of the accumulated meaning by the end of π_3, they remain contingent on the current monotonic content. As a result, when further monotonic content is added, such contingent inferences may change, without removing or editing any content directly.[2]

Nevertheless, as noted by Asher & Lascarides (2003), a dynamic logic can avoid this complexity altogether if one simply denies that discourse-pragmatic meaning is present in intermediate representations. Instead, it may be persistently underspecified, postponing the selection of any discourse-pragmatic interpretation until all other meaning has been fixed. In this way, persistent underspecification leaves no need for selective non-monotonicity in the representational system.

In this paper, we probe data from real-time comprehension to distinguish between these two possibilities for this flavor of discourse-pragmatic meaning. In particular, we search for evidence that intermediate representations include discourse-pragmatic meaning which must be revised in cases like (1). We rely on the assumption that during incremental comprehension, when a previously-formed representation becomes inappropriate, the comprehension mechanism suffers a detectable cost, what we will call REGRET. As we will review in §2, the psycholinguistic evidence for

1 For an overview of dynamic semantics, see Nouwen, Brasoveanu, van Eijck & Visser (2022).
2 Jaspars & Kameyama (1998) and Haug (2014) explore similar proposals.

REGRET in cases like (1) is thin and inconsistent, with most studies failing to observe the critical effect. We put an underspecification account to a stronger test than in this previous literature, examining heavily-biased and late-resolving cases of ambiguity as in (1), and we find clearer evidence for REGRET (§3–§5). This evidence motivates a theory of incremental interpretation that can account for the presence and revision of discourse-pragmatic meaning in at least some partial representations. A theory that adopts persistent underspecification throughout is therefore unsatisfying. We conclude in §6 by discussing how a dynamic semantic theory based on Lascarides & Asher (1993) and Asher & Lascarides (2003), together with some assumptions about real-time processing, could explain the results of this study, together with similar results from our research group (Duff & Altshuler 2024).

2 Previous work: No REGRET?

We turn to incremental processing data here armed with the assumption that the processing of new linguistic material should entail some momentary difficulty when it is incompatible with a previously-constructed representation or preference among representations. We call the source of this difficulty REGRET in order to encompass claims that this cost comes from the particular difficulty of reanalysis or revision to a representation or preference (Frazier & Rayner 1982; Huang, Arehalli, Kugemoto, Muxica, Prasad, Dillon & Linzen 2024), or the cost of misplaced lexical or structural expectations on the basis of the prior representation (Hale 2006; Levy 2008).[3]

In some prior work on semantic processing, researchers have looked for evidence of REGRET in order to distinguish between the questions of meaning which are resolved during initial comprehension—and can lead to REGRET—and those which may be left underspecified. For instance, Frazier & Rayner (1990) find that unexpected disambiguating material elicits apparent REGRET when it follows and disambiguates a homonym (*jam* as a condiment vs. an obstruction), but not when it does so for a polyseme (*newspaper* as a printed object vs. a corporation), supporting a hypothesis that only ambiguities of distinct grammatical representations are resolved within the sentence (see Frazier 1999; Frisson 2009 for further discussion).

Are the ambiguities of relative temporal order and causal relation in (1) represented or underspecified? In the remainder of this section, we review prior work on this question in incremental comprehension, highlighting why REGRET is the pattern

3 Under this way of categorizing sentence processing effects, we would classify well-known "garden path" costs (e.g., Frazier & Rayner 1982) as a robust REGRET effect at the level of syntactic representations. We use REGRET rather than "garden path" as our general term here in order to (a) foreground the core assumption, that this difficulty emerges when an initial representation turns out to be inappropriate given later input, and (b) avoid any associations that the term "garden path" may have with extreme, conscious difficulty, or reanalysis-based accounts specifically.

to look for, and recounting previous studies which did not observe this pattern.

We will follow much previous work in treating this ambiguity as one of *discourse coherence*, where temporal order and causal inference are consequences of establishing *coherence relations*, which implicitly structure the elements of a discourse.[4] We will assume that every new unit of the discourse π_b must participate in some relation with a previous (possibly complex) unit π_a—we'll call π_a the *head* and π_b the *tail*. In the case of (1), we can say there is an ambiguity between an initially-preferred relation Explanation(π_2, π_3), and a subsequently-preferred relation Result(π_2, π_3).

While there has been much research on the role of coherence in incremental comprehension, we lack robust evidence that would show us that ambiguities of coherence are resolved. One reason for this is that foundational work on discourse processing has largely focused on a separate question: Do effectively-unambiguous coherence relations marked by a connective (e.g., *because*) influence expectations for the form and content of the tail? Despite early debate (see Stewart, Pickering & Sanford 2000), the majority of modern studies has provided evidence in favor, from reading times (Koornneef & Van Berkum 2006; Mak & Sanders 2010; Canestrelli, Mak & Sanders 2013; Scholman, Rohde & Demberg 2017; Asr & Demberg 2020; Marchal, Scholman, Sanders & Demberg 2024a) and other measures of real-time expectation (Pyykkönen & Järvikivi 2010; Xu, Jiang & Zhou 2015; Xiang & Kuperberg 2015). But this pattern does not answer our question fully. If discourse meaning only found its way into incremental semantic representations when coherence was effectively disambiguated by cues like *because*,[5] this would not require the same machinery as assumed by theories of discourse coherence,[6] nor would it rule out the possibility that this layer of meaning remains persistently underspecified until fully disambiguated. In other words, this alone does not show us that representations must

4 For overviews of formal theories of discourse coherence, see, e.g., Zeevat (2011); Kehler (2019); Jasinskaja & Karagjosova (2020); Altshuler & Truswell (2022); Sasaki & Altshuler (2025).

5 On one hand, there is a tradition of treating connectives as explicit and unambiguous markers of coherence relations (e.g. Webber 2004), contrasted with 'implicit' coherence relations which lack a connective (see Zufferey & Degand 2024 for an overview). On the other hand, connectives like *because* have multiple meanings: e.g. *because* largely cues Explanation, but also sometimes has a *diagnostic* use relating some claim to its evidence (*It is raining because the streets are wet.*) (Traxler, Sanford, Aked & Moxey 1997; McHugh 2023). Therefore, it may be correct to treat connectives as just another indirect cue for a particular coherence relation. Nevertheless, we will treat connectives as a categorically privileged class of cues. We are comfortable with this simplification for a few reasons. For one, as reported by Traxler et al. (1997), there is incremental processing evidence that the default interpretation of *because* is causal (see also Solstad & Bott 2017). We imagine that such defaults hold for other connectives and their principal uses as well. And, in particular, in the narratives we will discuss, we observe that connectives cannot be overridden to the same extent as even the strongest implicit cues we will introduce. For instance, if (1) were to relate π_2 and π_3 using *because*, Explanation would be the only possible interpretation, even with the later content of π_4.

6 See Zufferey & Degand (2024) and Sasaki & Altshuler (2025) for recent discussion.

be subject to non-monotonic revision. In order to motivate this, we need to observe that discourse meaning is represented even while it remains ambiguous, so that it is vulnerable to bias-driven mistakes and REGRET.

A smaller array of previous work has sought signals of REGRET in order to provide evidence for incremental representations of discourse-pragmatic meaning, with variable results. The closest parallels to (1) come from Dickey (2001) and Sasaki (2021), who investigate similar discourses where implicit ambiguities of coherence travel hand in hand with ambiguities of temporal order. For instance, consider (2a), a stimulus from one of Sasaki's self-paced reading experiments. Before the bolded portion of π_2, comprehenders may prefer to interpret π_2 as describing a pinching event which followed the poking event in π_1, in retribution. If comprehenders registered this temporal order rapidly and firmly, one would expect to observe a momentary delay due to REGRET during the reading of the bold disambiguating region—compared to how it would be read in (2b), where it would come before any interpretation of π_2.[7] Interestingly, neither Sasaki nor Dickey observed any costs specific to this late disambiguation.

(2) Stimuli from Sasaki (2021), Experiment 1

a. [Lianne poked Eric.]$_{\pi_1}$ [He pinched her on the arm **a moment earlier**, in his seat near the back of the lecture hall.]$_{\pi_2}$

b. [Lianne poked Eric.]$_{\pi_1}$ [**A moment earlier**, in his seat near the back of the lecture hall, he pinched her on the arm.]$_{\pi_2}$

Mak & Sanders (2013) and Duff (2023) used similar manipulations to investigate the profile of cases where discourse coherence may lead to a causal inference which later provokes REGRET. For instance, consider (3a), a stimulus from one of Duff's experiments measuring reading times using the Maze task (Forster, Guerrera & Elliot 2009). Before the bolded *because*-clause, comprehenders may prefer to interpret the relative clause in π_1 as providing an explanation of Sally's choice (Cohen & Kehler 2021; Hoek, Rohde, Evers-Vermeul & Sanders 2021a). If readers register this causal inference rapidly, we would expect REGRET at the bold *because*-clause, because this clause provides an alternative explanation and encourages retraction of the initial inference. This REGRET would lead to slower reading times on the *because*-clause in (3a) when compared to (3b), where the *because*-clause comes before there is any inference to retract. However, no such delay was observed.[8]

7 To factor out the differences associated with the position of the disambiguating material, this movable-disambiguator experimental design always compares against a disambiguating region that pushes in the other direction. Key evidence for REGRET would be particular costs associated with late disambiguation in one direction.

8 The same lack of a retraction cost is true of a similar self-paced reading study by Mak & Sanders (2013) with causal inferences between sentences.

(3) Stimuli from Duff (2023), Experiment 8

Sally lives in a small city, where recently there was a citywide election for a new mayor with several candidates, and she had to decide among them on her mail-in ballot.

a. Last week, [she voted for the candidate that has a progressive platform, Pat Mirabella,]$_{\pi_1}$ [**because his name is first on this year's ballot**.]$_{\pi_2}$

b. Last week, [**because his name is first on this year's ballot**,]$_{\pi_2}$ [she voted for the candidate that has a progressive platform, Pat Mirabella.]$_{\pi_1}$

In sum, previous experiments testing for REGRET in the cancellation of causal inferences or the reversal of inferred temporal order have not found evidence that ambiguous discourse coherence is resolved in incremental representations. In contrast, some studies have observed REGRET associated with other ambiguities which coherence may influence, providing evidence that coherence-related processing can be rapid and incremental. For instance, Rohde, Levy & Kehler (2011) demonstrate that coherence-related pressures can influence the preferred attachment site of a relative clause. In typical cases where a relative clause follows a complex NP, English comprehenders prefer to interpret the relative clause as modifying the lower constituent. Indeed, plural agreement *are* in (4a) triggers apparent REGRET when it indicates that this relative clause is not modifying the lower constituent *the musician*, but instead the complete object, *the children of the musician*. Rohde and colleagues found that this REGRET could be eliminated when an implicit causality verb like *detests* set up an expectation for an upcoming explanation about the children's behavior, viz. (4b), suggesting that this expectation reduced the typical bias.

(4) Stimuli from Rohde et al. (2011), Experiment 2

a. John babysits the children of the musician who are generally arrogant...

b. John detests the children of the musician who are generally arrogant...

To explain this result, one would have to admit that discourse-based expectations are formed during processing; they must even be formed early enough to influence syntactic processing.[9] Nevertheless, this does not yet provide evidence that discourse meaning is itself represented, in such a way that it provokes REGRET. In order to account for this effect while explaining why REGRET does not surface in the other

9 Compare effects of referential context on initially-preferred syntactic analyses summarized in Steedman & Altmann (1989). In another strand of literature, researchers have debated about whether expected-but-still-ambiguous coherence relations can also influence lexical predictions. While there is some evidence in support of this (Kuperberg, Paczynski & Ditman 2011), other work has cast doubt on the strength, immediacy, and typicality of these effects (Murray 1997; Mulder 2008; Grisot & Blochowiak 2017; Crible & Pickering 2020; Marchal, Scholman & Demberg 2024b).

experiments reviewed above, Duff (2023) proposes that perhaps discourse meaning provokes gradient biases on other representations, like syntax, but it does not get represented itself in the same way. This state of affairs would be a theoretically unappealing compromise between these two classes of processing data, but it seems appropriate given the current evidence, and it would leave the door open for persistent underspecification.

In light of this challenging empirical terrain, the value of further work probing for discourse-specific REGRET is clear. To this end, we revisit here the types of discourses probed by Dickey (2001) and Sasaki (2021), where an ambiguity of implicit coherence is associated with a difference of temporal order between two adjacent clauses, in stimuli much like example (1). In order to provide the strongest possible test for REGRET, we constructed items with the intention of producing and then reversing the strongest possible coherence bias. If preference-reversing continuations remain REGRET-free even in these cases, persistently-underspecified coherence would remain plausible. Instead, as we will detail, these stronger cases reveal clear evidence for REGRET in real-time comprehension, and clarify our need for incremental and revisable discourse-pragmatic representations.

3 Experiment 1: Interpretation preferences

In an initial experiment[10] we collected preferred temporal order interpretations of 60 potentially-ambiguous narratives much like our initial motivating example (1), with and without disambiguating continuations. Each narrative featured a scene-setting π_1, followed by a sequence of simple past clauses (π_2, π_3) that affords two temporal orders: either a temporal progression where π_3 completes Result (i.e. cause/effect), or a temporal backshift where π_3 completes Explanation (i.e. effect/cause). π_2 contained an object implicit causality verb like *sued* (Garvey & Caramazza 1974), expected to bias towards Explanation, so long as π_3 discussed the object of π_2.

In order for these stimuli to serve as suitable probes for REGRET, they must provide the expected strong bias in favor of Explanation and a backshifted temporal order, and it must also be possible to later disambiguate against that interpretation. To verify, we elicited binary temporal order interpretations followed by 4-point judgments of certainty, across six key conditions. To assess the degree of bias present in the critical ambiguous Explanation-biased condition, we compare it to a 'de-biased' π_3 which was made to be an implausible cause via argument reversal, and to two unambiguous variants where π_3 began with a discourse connective marking Explanation (*because*) or Result (*so*). To verify that initial interpretations can change,

10 The materials, methods, and analytical choices for both experiments were pre-registered before data was collected for Experiment 1 (Duff, Sasaki & Altshuler 2025b). Data and analysis scripts are available in our supplementary materials (Duff, Sasaki & Altshuler 2025a).

Tensions were high in the shops on Main Street.

Temp. Order	Context	Text
Ambiguous	Expl-biased De-biased	Phil sued Liz. She defaced his shop on Tuesday, with red paint. Phil sued Liz. He defaced her shop on Tuesday, with red paint.
Disambig (Back/Prog)	Expl-biased	Phil sued Liz. She defaced his shop on Tuesday, with red paint, {so he decided to get lawyers involved/because his lawsuit was cruel}.
Unambig (Back/Prog)	Expl-biased	Phil sued Liz, {because/so} she defaced his shop on Tuesday, with red paint.

Given the way the story was told above, which event most likely happened first?
Phil sued Liz Liz defaced Phil's shop
How likely is the other option? (1 2 3 4)

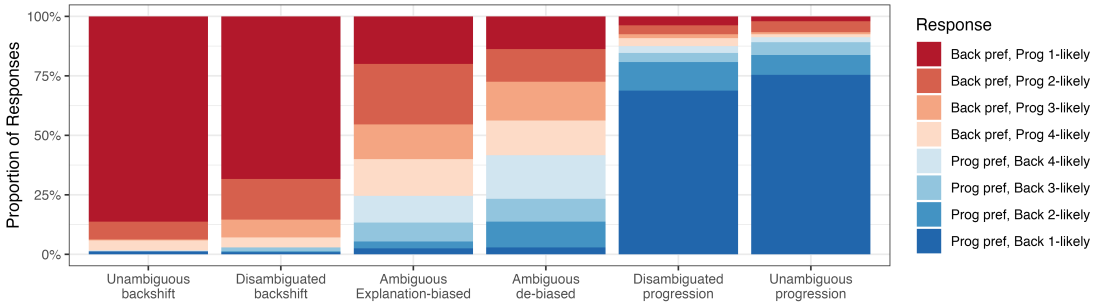

Figure 1 Stimuli and response distributions from the conditions of Experiment 1.

we also examine two conditions which continue with a π_4 which disambiguates temporal order by referring to a quality of π_2 as an explicit cause or result of π_3.

We tested 30 native English speakers recruited on Prolific (Figure 1). Binary order interpretations and certainty responses were combined to create a single index of the direction and strength of interpretation preferences,[11] which was then analyzed using Bayesian ordinal mixed-effects regression.[12] Participants exhibited a strong preference (75%) for a backshift in the ambiguous, Explanation-biased condition. This preference was higher than the de-biased condition, $\delta_{.95} = (-0.60, -0.13)$, and was fully reversed by an π_4 disambiguating to a Result, $\delta_{.95} = (-2.81, -1.93)$. These results replicate a strong backshift bias in such narratives, and further confirm

11 An 8-point ordinal variable ranging from temporal backshift preferences where a temporal progression was deemed unlikely, to progression preferences where an backshift was deemed unlikely.

12 For both experiments, analyses were conducted using `brms` (Bürkner 2017, 2018) in R, using regularizing priors. In the text, we report 95% highest-density intervals over the posterior marginal differences between conditions extracted from our models, $\delta_{.95}$. When this interval excludes 0, we take the associated effect to be noteworthy. δ in Experiment 1 is expressed in abstract model-internal units, not the units of the response scale; in Experiment 2, δ is the difference in reading times in milliseconds.

that preferred temporal orders can be reversed by a subsequent clause.

4 Experiment 2: Phrase-by-phrase reading times

Using the results of Experiment 1, we identified the 24 narratives with the strongest temporal backshift bias in π_3 and the largest reversal given π_4,[13] to investigate in a chunked, non-cumulative self-paced-reading experiment testing for REGRET. We compared these 24 items across three conditions (see top of Figure 2). In (initially) Ambiguous conditions, the sequence is expected to prefer Explanation, but it is followed by a *because*-clause π_4 which requires Result, as tested in Experiment 1. In both other conditions, the same π_4 would not provoke REGRET against an initial interpretation, because π_3 already makes a Result unambiguous given the adverb *in response*. If comprehenders indeed interpret the coherence of π_2 and π_3 before π_4, we should see particularly slow reading of π_4 in the Ambiguous condition.

Additional investigations of reading times on the adverbial cue *in response* in π_3 allow us to diagnose the possibility for earlier coherence interpretation triggering earlier REGRET. In our second condition (Progression Adverb), *in response* is the first cue against Explanation and the temporal backshift. By the same logic as above, this region could demonstrate REGRET if a discourse interpretation has already been computed before the offset of π_3. We can observe this cost by comparing against the final condition (*so* + Progr. Adv.), where *so* forces the comprehender towards the Result reading from the beginning of π_3.

90 participants (recruited as in Experiment 1) read these narratives in chunked, non-cumulative self-paced reading format, mixed with 56 fillers.[14] After each narrative, participants answered a two-alternative forced-choice question, with subject matter distributed equally across various parts of the text. For instance, for the item in Figure 2, participants were asked about the scene-setting information in π_1: "What were these characters?" ("Business owners," or "Landlords"?) In six of the 24 critical trials, questions asked about the order of the events reported in π_2 and π_3. These questions were included to motivate participants to comprehend of all aspects of the narratives, including (but without any special bias towards) temporal order.

13 To establish this ranking, we extracted item-by-item predicted temporal backshift biases from our mixed-effects model in the biased ambiguous condition, and the disambiguated backshift condition. Model predictions in the ambiguous condition were taken to measure backshift bias, and the difference in model predictions between the two conditions was taken to measure reversal strength. These two indices were z-scored and summed to provide an overall measure of item quality.

14 Fillers included 24 narratives taken from the original set of 60 above, shown with their Explanation-biasing π_4. Participants saw half of these fillers in which π_3 was introduced as an explicit Explanation using *because*, and half where the coherence relation was left ambiguous until π_4. The remaining 32 fillers were of a similar length, using a variety of clausal connectives, with half featuring ambiguous pronouns late in the narrative.

Tensions were high/in the shops/on Main Street.	
Ambiguous	Phil sued Liz./She defaced his shop/on Tuesday,/**with red paint,**
Progression Adv.	Phil sued Liz./She defaced his shop/in response,/**with red paint,**
so + Progr. Adv.	Phil sued Liz,/so she defaced his shop/in response,/**with red paint,**
because his lawsuit/**was cruel**. **She was**/ready to destroy his reputation/in the town.	

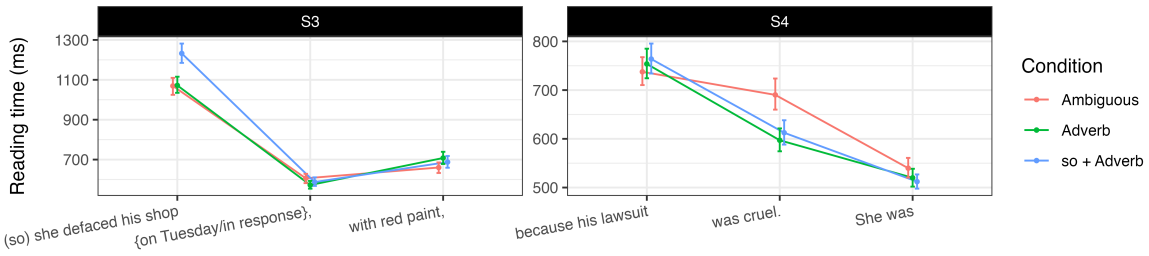

Figure 2 Stimuli and reading times across the critical regions of Experiment 2. "/" indicates a chunk boundary. Critical regions in bold.

Reading times in the critical items (Figure 2) were analyzed using Bayesian log-normal mixed-effects regressions. As for our main prediction, we find that π_4 was indeed read slowest in the Ambiguous condition (red), with credible slowdowns of about 60ms at the end of the disambiguating *because* clause, $\delta_{.95}$ = (34ms, 82ms), and 20ms on the subsequent chunk at the beginning of π_5, $\delta_{.95}$ = (9ms, 36ms). We take this to be reflective of REGRET associated with a representation of the coherence structure for the preceding discourse units.

Examining reading times at *in response* between conditions, we find evidence that this cost is unique to a later disambiguation. Comparing the Adverb condition (green) to the *so* + Adverb condition (blue), we lack strong evidence for any slowdown on *in response* where it served as the first disambiguation to Result, $\delta_{.95}$ = (-26ms, 5ms), nor any slowdown on the following spillover region $\delta_{.95}$ = (-6ms, 39ms). In other words, there was no REGRET-like cost for a late cue within π_3. Our only evidence for REGRET comes from the later disambiguation within π_4.

5 Discussion of experimental findings

In Experiment 2, we observed a slowdown during the incremental processing of temporarily ambiguous discourses. We take this slowdown to be best explained as REGRET following representation of coherence and temporal order. In this section, we discuss how this effect compares to other processing evidence, and weigh some alternative explanations, before moving on in §6 to discuss the implications for representational theories.

5.1 Ruling out an alternative explanation

Another plausible explanation comes to mind: this slowdown is a function merely of interpreting a dispreferred coherence relation. This alternative would suffice to explain the effect without assuming a representation which must be revised—but we review here why we do not credit it in this case.

Work in sentence processing has indeed provided evidence that certain representations are difficult to construct in certain contexts, not because of the presence of a prior representation as in REGRET, but because there is a conflict between a preferred representation in isolation, and the representation that it must have in context. We will call such effects instances of TENSION, which are known to arise e.g. in the processing of homonyms. In the REGRET effect discussed in §2, difficulty arises when disambiguating information follows a homonym, and obligates the dispreferred interpretation of the homonym. But when a homonym occurs in a context which already requires a dispreferred meaning, researchers have observed a slowdown at the homonym itself, attributed to a difficult process of initial interpretation, as the comprehender manages conflicting sources of preference (Duffy, Morris & Rayner 1988; Kambe, Rayner & Duffy 2001). In all our critical stimuli, there is also a conflict between two sources of bias, (a) the implicit causality verb in π_2 raising interest in an Explanation, and (b) the various disambiguating cues that we presented towards the alternative Result interpretation. Costs on any of these disambiguating cues could be seen as TENSION during the initial selection of a coherence relation.

Nevertheless, our experimental design allows us to isolate the expected costs of TENSION in this case, and those costs on their own cannot explain the effect on π_4. We can isolate TENSION by comparing the profile of π_3 in our first condition—where no alternative cue has yet been encountered—and in our latter conditions—where some disambiguating cues (*so* and/or *in response*) are already present, and can trigger selection of a coherence relation. Extra slowdown in the latter case can be attributed to the degree of TENSION between the disambiguation cues and prior bias from π_2. The first or second regions of π_3 also differ in their lexical material, but we can compare reading times in the final region, which was held constant. This final region does suggest a slowdown in the conditions featuring *in response* (blue and green in Figure 2), $\delta_{.95}$ = (1ms, 41ms).[15] This slowdown, however, is quite small when compared to the ∼60ms slowdown present in π_4 in the Ambiguous condition. As a result, we consider it unlikely that these are two equal occurrences of TENSION

15 Sasaki (2021) (Experiment 1B) observes a similar possibility when an explicit progression adverbial ends a plausible Explanation tail, while Dickey (2001) finds the opposite in his materials, a slowdown associated with marking a backshift. Given that Dickey's materials seem to prefer progression, it is easiest to explain these costs as TENSION against the particular context, and not some default preference for temporal progression or backshift. See Mak & Sanders (2013), Hoek, Rohde, Evers-Vermeul & Sanders (2021b), and Duff (2023) for similar findings in causal inference.

that are associated with the parsing of a Result interpretation. TENSION may explain a portion of the later difficulty in the Ambiguous condition, but there remains some additional slowdown which seems to be particular to a later disambiguation. This additional slowdown is nicely explained by REGRET.

5.2 Converging evidence from coherence-related pronominal ambiguities

In other recent work along similar lines (Duff & Altshuler 2024), our research group searched for evidence of REGRET in the processing of another discourse-coherence ambiguity, correlated with an ambiguity of pronominal coreference.[16] Much of this work has demonstrated REGRET associated with pronominal coreference, using various ways of creating an initial bias. Costs have been observed when a pronominal antecedent must be revised contrary to a prominent referent highlighted across many previous clauses (Gordon & Scearce 1995); contrary to the bias from an explicitly-established coherence relation (as discussed above, e.g. Koornneef & Van Berkum 2006); contrary to the bias from an explicitly-provided question-under-discussion (Kehler & Rohde 2017); and contrary to a bias from world knowledge (Jones & Bergen 2024). But none of this previous work investigated cases of ambiguous coherence, where a strong bias for a particular implicit coherence relation might be driving a preferred interpretation of the pronoun. Much like the case we reported here, this could serve as evidence that discourse meaning is actively constructed during incremental interpretation, even when it is not disambiguated in the input.

Duff & Altshuler (2024), henceforth D&A, sought to address this gap in evidence, targeting a joint ambiguity of discourse coherence and pronominal antecedence where preferred interpretations were well-established in previous work. In cases like (5a), the object pronoun at the end of π_2 is typically interpreted as coreferent with the object of π_1, here *Winona* (Smyth 1994). Prior research has demonstrated that this is not merely due to syntactic parallelism, but a contingent decision to relate π_1 and π_2 via a Parallel relation; once another relation is marked or inferred, the most frequent interpretation of the pronoun shifts to *Mia* (Wolf, Gibson & Desmet 2004; Kehler, Kertz, Rohde & Elman 2008). D&A constructed and normed such cases so that they featured a strong initial bias for Parallel. They then measured reading times in chunked, non-cumulative self-paced reading when these biased narratives were followed by a π_3 as in (5b), where the presupposition of *in return* required Ian to have kicked Mia, and a Result interpretation becomes salient.

(5) a. [Mia hit Winona with a pea]$_{\pi_1}$ and [Ian kicked her under the table.]$_{\pi_2}$

16 Pronominal ambiguities have been a perennial focus of work on discourse processing (more than temporal ambiguities discussed above). In addition to the work cited here, see, e.g. Chambers & Smyth (1998); Kertz, Kehler & Elman (2006); Stevenson, Knott, Oberlander & McDonald (2010); Kaiser (2011); Kaiser & Cherqaoui (2016); Mozuraitis & Heller (2017).

b. [Mia didn't kick him in return...]$_{\pi_3}$

If a Parallel interpretation had already been represented, which preferred *Winona* as the antecedent, reading *in return* should elicit REGRET, when compared to conditions where the ambiguity had been resolved earlier (via gender on the critical pronoun, or via explicit discourse connective *so*). Indeed, D&A observed a 15ms slowdown on *in return,* and its subsequent spillover region. They take this slowdown, like ours, as REGRET brought about by the incremental analysis of discourse coherence.

5.3 Scheduled specification

One of the most interesting generalizations emerging from these findings is the evidence for the particular time course of coherence representation. It seems from the patterns observed here that the representation of coherence between two discourse units does not become settled in such a way that it could provoke REGRET until the end of the *tail* (the latter unit). In fact, the results from D&A align in this regard: REGRET was only observed on the first cue towards Result after the tail. In neither case did unexpected disambiguating material within the tail show any cost associated with being the first evidence against a preferred meaning.

As it happens, this timing profile is probably not unique to coherence information. Other ambiguities seem to be at first underspecified and only later represented. In particular, late disambiguation does seem to trigger REGRET for the representation of a polysemous noun if this disambiguation comes after a sentence boundary (Frisson & Frazier 2004 cited in Frisson 2009; Foraker & Murphy 2012), much as is the case here. In other cases, patterns of REGRET are variable as if representation is timed flexibly across contexts, tasks, and languages—for instance, see profiles of variable underspecification in the processing of ambiguous lexical aspect by Pickering, McElree, Frisson, Chen & Traxler (2006) and Bott & Hamm (2014).

This flexibility may help explain why we find evidence for discourse-oriented REGRET in these experiments, where others have not. If implicit coherence relations are typically represented only after the offset of the tail, it is no surprise that Dickey (2001) and Sasaki (2021) failed to find REGRET signatures at disambiguation points within the tail. Experiments by Mak & Sanders (2013) and Duff (2023) did look for REGRET through disambiguation after the tail, but they may have failed to achieve robust bias or motivate attentive comprehension—under those conditions, comprehenders may underspecify coherence for an atypically long time.

5.4 REGRET for coherence, or for its consequences?

At the end of §2, we suggested that results from Rohde et al. (2011) can be explained without REGRET over the representation of coherence, but instead simply as REGRET

concerning syntactic representations downstream of a less rigid treatment of discourse meaning. In the same way, it remains possible that the REGRET found in D&A comes about only because pronominal antecedence is represented and then revised, and the REGRET discussed here comes about only because the temporal order of two clauses is represented and then revised. Indeed, the evidence for distinctions of coherence almost always comes from downstream interpretive consequences, the specifications and inferences that shift together with the choice of a coherence relation. As a result, we could probably never isolate the costs of coherence-related REGRET from REGRET over these interpretive correlates.

This same problem shape arises in sentence processing: evidence of REGRET does not prove straightforwardly that syntactic structures are being represented and revised incrementally, because the syntactic ambiguities under investigation have their own interpretive correlates as well. Nevertheless, even when researchers highlight the role of specific changes in meaning in producing REGRET (e.g. the focus on thematic role identification in Pritchett 1988, Frazier & Clifton 1996, and Sturt & Crocker 1997), they do not give up the underlying hypothesis that these meanings are determined by incrementally-composed syntactic parses. On our read of the field, this is because the evidence elsewhere for syntactic structure as a component of linguistic competence is so strong, and because the alternative (deriving interpretations which are equally constrained without reference to an underlying abstract representation) is so theoretically unappealing.

On the basis of similar support, we prefer positing incrementally-built coherence representations to whatever representation-free alternative one could imagine. Theories of coherence offer a way to capture the relationships between disparate elements of discourse meaning, and now that we have wider evidence that many of those disparate elements can be associated with REGRET, the most appealing explanation is that coherence representations are constructed in real-time comprehension.

6 Consequences for representational theories

The results from Experiment 2 contribute to mounting evidence that incremental comprehenders represent and revise discourse meaning. In the face of this evidence, insofar as dynamic semantic representations seek to model the mental representations language users actually employ during discourse comprehension, these representations must be capable of selecting, and not persistently underspecifying, discourse meaning, and they must admit some mechanism for non-monotonic update.

That being said, the exact consequences here depend on how one chooses to distribute the burden of explanation between a theory of representation and a theory of incremental processing (Brasoveanu & Dotlačil 2015). We conclude by outlining how such consequences could take shape depending on this division of labor.

6.1 Representations permitting non-monotonic edits and underspecification

One option is to design the representational system with an eye towards processing; in this case, to define a dynamic semantics which permits all and only the states and state transitions used by human incremental comprehenders. For theories taking this route, our results suggest that this semantics must be capable of both representing and temporarily underspecifying discourse meaning, and that those particular discourse representations must permit selective non-monotonic update.

The version of Segmented Discourse Representation Theory (SDRT) laid out by Asher & Lascarides (2003) meets these criteria nicely, with discourse meaning calculated through application of violable constraints to the output of monotonic interpretation. Calculating the relations and other specifications which optimize against those constraints can be flexibly timed. Once this calculation is performed, it provides representational choices which can be frozen (thereby representing coherence), but it can also always be performed again at any point in interpretation without retracting monotonic content (thereby permitting selective non-monotonicity).

This approach parses our critical example (1), shortened here as (6), as follows.

(6) [Stella demoted Gary.]$_{\pi_1}$ [He skipped a meeting,]$_{\pi_2}$ [because the demotion angered him.]$_{\pi_3}$

In SDRT, update reduces a set of representations of discourse meaning σ. Reductions come about as the interpreter adds hard constraints that capture the monotonic meaning of each new discourse unit. At any point, the members of σ are thus all of the possible representations consistent with the discourse so far. A separate system, Glue Logic,[17] ranks the members of σ by generating defeasible inferences based on the monotonic content and other soft factors (a principle of maximizing coherence, a variety of axioms for inferring particular coherence relations, etc.).

Beginning (1), all members of σ include a discourse unit π_1, which describes an event e_1 of Stella demoting Gary. Continuing, the monotonic content of π_2 fixes an event e_2 of some individual x skipping a meeting. Glue Logic then introduces other defeasible inferences via ranking. Some of these are trivial, like the assumption that x is Gary. Because world knowledge strongly suggests that skipped meetings can cause demotions, Glue Logic will also rank as highest those representations which include Explanation(π_1, π_2)—and this relation requires that the event described by π_2 caused (and therefore preceded) the event described by π_1. At this point, representations which include Result(π_1, π_2) also remain in σ, but are ranked lower given the relatively weaker possibility that the demotion caused the skipped meeting.[18]

17 See Lascarides & Asher (2007); Altshuler & Schlöder (2019) for an overview of Glue Logic.

18 Multiple relations may hold between a pair of discourse units (π_1, π_2). Nevertheless, representations including both Explanation(π_1, π_2) and Result(π_1, π_2) are blocked because they impose conflicting

If σ were merely an unranked set, this output would be underspecified for coherence, as multiple interpretations remain in the candidate set. However, Glue Logic produces an incrementally-determined rank over these candidates. This output rank, under common linking hypotheses in theories of processing, would be expected to generate REGRET much in the same way as a single analysis. One possibility is that this rank guides incremental expectations for future forms or content (Levy 2008), rendering continuations that are incompatible with top-ranked analyses difficult to integrate. Another possibility is that changes to the inference preferences which determine the output rank are themselves costly. A final option, closely related to other proposals in sentence processing (Jurafsky 1996; Boston, Hale, Vasishth & Kliegl 2011; Huang et al. 2024), is that σ only retains a limited number of minimal analyses of previous context, with well-supported inferences becoming frozen over time, categorically pruning away the representations they disprefer.[19]

When the interpreter comes to the monotonic content of π_3, it adds the restriction that there must be a discourse unit π_3, which describes an event e_3 wherein some already-established demotion event d angers some individual y, and this described event e_3 must have caused the event described by π_3's syntactic host π_2. This last restriction is an unavoidable cue towards Explanation(π_2, π_3) in Glue Logic, and thereby the event described by π_3 must have preceded that of π_2. Meanwhile, the only apparent antecedent for the definite demotion d is the event e_1 from π_1. Because of this coreference, candidates extending the previously preferred candidate cannot prevail here: if the skipping e_2 preceded the demotion e_1, the demotion e_1 could not also have participated in an even-earlier angering e_3. The final interpretations will give up on the previous inference of Explanation(π_1, π_2), in favor of Result(π_1, π_2) and Explanation(π_2, π_3). Whether because of poor predictions, because of a shift in the inferences delivered by Glue Logic, or because candidates featuring Result(π_1, π_2) were pruned from σ, this last step would be associated with REGRET.

Note that as the monotonic content of π_2 unfolded in time, it would in principle be possible to run Glue Logic ranking on partial evidence, and still end up positing Explanation. If this happened with some frequency, we would have expected that the late adverbial within π_2 would have also triggered REGRET. We can account for this if this computation is not typically run until the sentence boundary.

requirements on the discourse structure, and contradictory entailments about time and causality (though see Copley & Martin 2014). For more on Result, see Andersson & Spenader (2014); Altshuler & Varasdi (2015); for more on Explanation and Result comparison, see Altshuler (2021).

19 In a pruning approach, the interpreter may sometimes end up with suboptimal parses because the globally optimal candidate was pruned at an earlier stage. To our knowledge, this prediction has never been investigated in coherence, and deserves future study.

6.2 Simpler representational approaches

We considered above how an articulated dynamic theory could treat incremental representation and revision of coherence in line with the processing patterns we observe. But this is not a necessary conclusion: at another extreme, one might adopt a radically simple representational theory, shifting the explanatory burden of incrementality onto a more complex processing theory. And, between these extremes, one could adopt some incremental complexity into the representations, as in a traditional dynamic semantics, but still leave the finer details of building and editing these representations to the processing mechanism; Brasoveanu & Dotlačil (2020) offer a good starting place for this kind of account.

D&A provide an argument against the simplest representational theories by observing that the slowdowns they attribute to discourse-level REGRET are much smaller (15ms) than the effects attributed to REGRET in syntactic parsing (50ms in Ferreira & Henderson 1990). One way to explain this would be if syntactic REGRET involved destructive edits to the current parse state, while discourse REGRET may require only some more accessible recomputation. For this to hold, we'd want a representational system which at least distinguished destructive edits and accessible recomputations. This conclusion, however, is already somewhat less clear given the 50ms slowdown we attribute to discourse REGRET in Experiment 2. Nevertheless, these kind of comparisons offer a promising tool for narrowing the space of plausible theories. We hope future experiments can shed further light on whether this REGRET is categorically different from the costs observed in experiments on other phenomena.

6.3 Comparing REGRET: A new argument for discourse structure?

On that last note, the slowdown in this study was many times larger than the slowdown in D&A. Could this difference be meaningful?

We advise caution, given that participant and item differences stymie direct comparison. Nevertheless, there are notable differences between the interpretive adjustments triggered across the two studies. One such difference is in hierarchical discourse structure: in SDRT, choices of coherence relations not only introduce associated inferences, but also control the accessibility of previous discourse units (Asher & Vieu 2005).[20] *Subordinating* relations like Explanation preserve access to their heads, while *coordinating* relations like Result or Parallel block it.[21] The

20 This is also true in other theories of discourse structure; see e.g., Hobbs 1985; Polanyi 1985; Grosz & Sidner 1986; Mann & Thompson 1986; van Kuppevelt 1995; Webber, Stone, Joshi & Knott 2003, a.o.

21 This is because coordinating relations change the scene, and hence *move forward* the narrative, while subordinating relations detail the scene, and thereby *deepen* the narrative (Altshuler & Schlöder 2019); see Jasinskaja & Karagjosova (2020); Hunter & Thompson (2022) for more discussion.

hypothetical revision from Parallel to Result in D&A has no effect on this aspect of the discourse, but the revision here from Explanation to Result does. The slowdown produced by REGRET may be proportional to the difference between representations; if so, we might understand this extra slowdown as evidence for the cognitive reality of the coordinating/subordinating distinction.[22]

That being said, other differences abound. Another attribute which seems to make the shift from Explanation to Result special is the presence of a reversal in both causal structure and temporal order, where D&A's shift from Parallel to Result only introduces a new causal inference. Perhaps the extra penalty, then, is because this reversal of causal inferences[23] is especially complicated to manage.[24]

As we continue expanding this method to probe other discourse ambiguities, we are currently planning for our next experiments to discriminate between these two possibilities for the gradations in REGRET, by probing shifts from subordinating to coordinating structure which do not involve causal reasoning nor require any inferential reversals. As an example of what we have in mind, consider (7):

(7) [The city council was discussing improvements to public transit.]$_{\pi_1}$ [They held an open forum about some proposals.]$_{\pi_2}$ [They debated a new subway,]$_{\pi_3}$ [because after the open forum they realized they had underestimated demand.]$_{\pi_4}$

Here, the initially-preferred interpretation takes π_3's debate to be a subpart of π_2's open forum, relating the segments by (subordinating) Elaboration. An alternative analysis, using (coordinating) Narration, is expected to be initially dispreferred, as it captures no relation between the open forum and the debate. π_4 nevertheless obligates this reading, placing the debate in the aftermath of the forum. If our larger REGRET slowdown in the present experiment can be explained as a consequence of structural differences, we expect a larger slowdown here too. On the other hand, if the present slowdown was inflated by the difficulty of inferential reversal, these results should be closer in size to the smaller slowdown in D&A.

22 Even though Parallel and Result are both coordinating relations in SDRT, they do differ in that the former is a structural relation requiring a so-called *common theme* from its arguments (see, e.g., Altshuler & Truswell 2022: Ch.6). This difference does not, however, appear to elevate REGRET, at least not to the same extent as the difference between subordinating and coordinating relations.

23 For more discussion of causal reversal, see, e.g., Sasaki (2021) and references therein.

24 Other differences include (a) The nature of the revision: Parallel is a gradable relation in SDRT; a weakened Parallel which does not require object identity may persist alongside Result in the final interpretation of D&A's stimuli. The full retraction of Explanation here may have provoked greater difficulty. (b) The nature of the cues: While measured biases were roughly equivalent here and in D&A, our bias was introduced by a single cue, an implicit causality verb, thought to be quite potent. The bias in D&A, arising from more distributed information, may have been easier to override.

References

Altshuler, Daniel. 2021. A puzzle about narrative progression and causal reasoning. In Andreas Stokke & Emar Maier (eds.), *The Language of Fiction*, Oxford University Press. doi:10.1093/oso/9780198846376.003.0010.

Altshuler, Daniel & Julian Schlöder. 2019. Anaphora and ambiguity in narratives. Lecture notes from ESSLLI 2019, University of Latvia, http://jjsch.github.io/teaching/esslli2019.html.

Altshuler, Daniel & Robert Truswell. 2022. *Coordination and the Syntax-Discourse Interface*. Oxford University Press. doi:10.1093/oso/9780198804239.001.0001.

Altshuler, Daniel & Károly Varasdi. 2015. An argument for definitional adequacy of RESULT and NARRATION. *Semantics and Linguistic Theory (SALT)* 25. 38–56. doi:10.3765/salt.v25i0.3049.

Andersson, Marta & Jennifer Spenader. 2014. RESULT and PURPOSE relations with and without 'so'. *Lingua* 148. 1–27. doi:10.1016/j.lingua.2014.05.001.

Asher, Nicholas & Alex Lascarides. 2003. *Logics of Conversation*. Cambridge.

Asher, Nicholas & Laure Vieu. 2005. Subordinating and coordinating discourse relations. *Lingua* 591–610. doi:10.1016/j.lingua.2003.09.017.

Asr, Fatemeh & Vera Demberg. 2020. Interpretation of discourse connectives is probabilistic: Evidence from the study of 'but' and 'although'. *Discourse Processes* 57(4). 376–399. doi:10.1080/0163853X.2019.1700760.

Boston, Marisa Ferrara, John T. Hale, Shravan Vasishth & Reinhold Kliegl. 2011. Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes* 26(3). 301–349. doi:10.1080/01690965.2010.492228.

Bott, Oliver & Fritz Hamm. 2014. Cross-linguistic variation in the processing of aspect. In *Studies in Theoretical Psycholinguistics*, 83–109. Springer. doi:10.1007/978-3-319-05675-3_4.

Brasoveanu, Adrian & Jakub Dotlačil. 2015. Incremental and predictive interpretation: Experimental evidence and possible accounts. *Semantics and Linguistic Theory (SALT)* 25. 57–81. doi:10.3765/salt.v25i0.3047.

Brasoveanu, Adrian & Jakub Dotlačil. 2020. *Computational Cognitive Modeling and Linguistic Theory*. Springer. doi:10.1007/978-3-030-31846-8.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80. 1–28. doi:10.18637/jss.v080.i01.

Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* 10(1). 395–411. doi:10.32614/rj-2018-017.

Canestrelli, Anneloes R., Willem M. Mak & Ted J. M. Sanders. 2013. Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes* 28(9). 1394–1413. doi:10.1080/01690965.2012.685885.

Chambers, Craig & Ron Smyth. 1998. Structural parallelism and discourse coherence: A test of centering theory. *Journal of Memory and Language* 39. 593–608. doi:10.1006/jmla.1998.2575.

Cohen, Jonathan & Andrew Kehler. 2021. Conversational eliciture. *Philosophers' Imprint* 21(12). doi:2027/spo.3521354.0021.012.

Copley, Bridget & Fabienne Martin. 2014. *Causation in Grammatical Structures*. Oxford University Press.

Crible, Ludivine & Martin J. Pickering. 2020. Compensating for processing difficulty in discourse: Effect of parallelism in contrastive relations. *Discourse Processes* 57(10). 862–879. doi:10.1080/0163853X.2020.1813493.

Dickey, Michael Walsh. 2001. *The Processing of Tense*. Kluwer.

Duff, John. 2023. *On the Timing of Decisions about Meaning during Incremental Comprehension*: UC Santa Cruz PhD dissertation.

Duff, John & Daniel Altshuler. 2024. Reanalysis in discourse comprehension: Evidence from reading times. *Amsterdam Colloquium* 24. 102–110. https://platform.openjournals.nl/PAC/article/view/21826.

Duff, John, Kelsey Sasaki & Daniel Altshuler. 2025a. Incremental interpretation of discourse coherence: Supplementary materials. doi:10.17605/OSF.IO/C32K6. OSF Repository.

Duff, John, Kelsey Sasaki & Daniel Altshuler. 2025b. Probing explanation-to-result reanalysis. doi:10.17605/OSF.IO/FVKC7. OSF Preregistration.

Duffy, Susan A., Robin K. Morris & Keith Rayner. 1988. Lexical ambiguity and fixation times in reading. *Journal of Memory and Language* 27. 429–446. doi:10.1016/0749-596X(88)90066-6.

Ferreira, Fernanda & John M. Henderson. 1990. Use of verb expectation in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16(4). 555–568. doi:10.1037/0278-7393.16.4.555.

Foraker, Stephani & Gregory L. Murphy. 2012. Polysemy in sentence comprehension: Effects of meaning dominance. *Journal of Memory and Language* 67. 407–425. doi:10.1016/j.jml.2012.07.010.

Forster, Kenneth I., Christine Guerrera & Lisa Elliot. 2009. The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods* 41(1). 163–171. doi:10.3758/BRM.41.1.163.

Frazier, Lyn. 1999. *On Sentence Interpretation*. Springer. doi:10.1007/978-94-011-4599-2.

Frazier, Lyn & Charles Clifton, Jr. 1996. *Construal*. MIT Press.

Frazier, Lyn & Keith Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14(2). 178–210. doi:10.1016/0010-

0285(82)90008-1.

Frazier, Lyn & Keith Rayner. 1990. Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language* 29. 181–200. doi:10.1016/0749-596X(90)90071-7.

Frisson, Steven. 2009. Semantic underspecification in language processing. *Language and Linguistics Compass* 3(1). 111–127. doi:10.1111/j.1749-818X.2008.00104.x.

Frisson, Steven & Lyn Frazier. 2004. Processing polysemy: Making sense of sense. Poster at CUNY 17, University of Maryland.

Garvey, Catherine & Alfonso Caramazza. 1974. Implicit causality in verbs. *Linguistic Inquiry* 5(3). 459–464.

Gordon, Peter C. & Kimberly A. Scearce. 1995. Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory & Cognition* 23(3). 313–323. doi:10.3758/BF03197233.

Grisot, Cristina & Joanna Blochowiak. 2017. Temporal connectives and verbal tenses as processing instructions: Evidence from French. *Pragmatics & Cognition* 24(3). 404–440. doi:10.1075/pc.17009.gri.

Grosz, Barbara & Candace L Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3). 175–204.

Hale, John. 2006. Uncertainty about the rest of the sentence. *Cognitive Science* 30. 643–672. doi:10.1207/s15516709cog0000_64.

Haug, Dag T. T. 2014. Partial dynamic semantics for anaphora: Compositionality without syntactic coindexation. *Journal of Semantics* 31. 457–511. doi:10.1093/jos/fft008.

Hobbs, Jerry. 1985. On the coherence and structure of discourse. Tech. Rep. CSLI–85–37 Center for the Study of Language and Information.

Hoek, Jet, Hannah Rohde, Jacqueline Evers-Vermeul & Ted J. M. Sanders. 2021a. Expectations from relative clauses: Real-time coherence updates in discourse processing. *Cognition* 210. 104581. doi:10.1016/j.cognition.2020.104581.

Hoek, Jet, Hannah Rohde, Jacqueline Evers-Vermeul & Ted J. M. Sanders. 2021b. Scolding the child who threw the scissors: Shaping discourse expectations by restricting referents. *Language, Cognition and Neuroscience* 36(3). 382–399. doi:10.1080/23273798.2020.1852292.

Huang, Kuan-Jung, Sutas Arehalli, Mari Kugemoto, Christian Muxica, Grusha Prasad, Brian Dillon & Tal Linzen. 2024. Large-scale benchmark yields no evidence that language model surprisal explains syntactic disambiguation difficulty. *Journal of Memory and Language* 137. 104510. doi:10.1016/j.jml.2024.104510.

Hunter, Julie & Kate Thompson. 2022. On the role of relations and structure in discourse interpretation. In Daniel Altshuler (ed.), *Linguistics Meets Philosophy*, Cambridge University Press. doi:10.1017/9781108766401.009.

Jasinskaja, Katja & Elena Karagjosova. 2020. Rhetorical relations. In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann & Thomas Ede Zimmermann (eds.), *Companion to Semantics*, Wiley.

Jaspars, Jan & Megumi Kameyama. 1998. Discourse preferences in dynamic logic. In Atocha Aliseda, Rob van Glabbeek & Dag Westertahl (eds.), *Computing Natural Language*, 67–96. Center for the Study of Language and Information (CSLI).

Jones, Cameron R. & Benjamin Bergen. 2024. Does word knowledge account for the effect of world knowledge on pronoun interpretation? *Language and Cognition* 16(4). 1182–1213. doi:10.1017/langcog.2024.2.

Jurafsky, Daniel. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* 20(2). 137–194. doi:10.1207/s15516709cog2002_1.

Kaiser, Elsi. 2011. On the relation between coherence relations and anaphoric demonstratives in German. *Sinn und Bedeutung* 15. 337–51.

Kaiser, Elsi & Boutaina Cherqaoui. 2016. Effects of coherence on anaphor resolution, and vice versa: Evidence from French personal pronouns and anaphoric demonstratives. In Anke Holler & Katja Suckow (eds.), *Empirical Perspectives on Anaphora Resolution*, 51–78. De Gruyter.

Kambe, Gretchen, Keith Rayner & Susan A. Duffy. 2001. Global context effects on processing lexically ambiguous words: Evidence from eye fixations. *Memory & Cognition* 29(2). 363–372. doi:10.3758/BF03194931.

Kehler, Andrew. 2019. Coherence relations. In Robert Truswell (ed.), *The Oxford Handbook of Event Structure*, 583–604. Oxford: Oxford University Press.

Kehler, Andrew, Laura Kertz, Hannah Rohde & Jeffrey L. Elman. 2008. Coherence and coreference revisited. *Journal of Semantics* 25(1). 1–44. doi:10.1093/jos/ffm018.

Kehler, Andrew & Hannah Rohde. 2017. Evaluating an expectation-driven Question-Under-Discussion model of discourse interpretation. *Discourse Processes* 54(3). 219–238. doi:10.1080/0163853x.2016.1169069.

Kertz, Laura, Andrew Kehler & Jeffrey L. Elman. 2006. Grammatical and coherence based factors in pronoun interpretation. *Annual Meeting of the Cognitive Science Society* 28. 1605–10.

Koornneef, Arnout W. & Jos J. A. Van Berkum. 2006. On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language* 54. 445–465. doi:10.1016/j.jml.2005.12.003.

Kuperberg, Gina R., Martin Paczynski & Tali Ditman. 2011. Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience* 23(5). 1230–1246. doi:10.1162/jocn.2010.21452.

van Kuppevelt, Jan. 1995. Main structure and side structure in discourse. *Linguistics* 33. 809–833. doi:10.1515/ling.1995.33.4.809.

Lascarides, Alex & Nicholas Asher. 1993. Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy* 16. 437–493. doi:10.1007/BF00986208.

Lascarides, Alex & Nicholas Asher. 2007. Segmented Discourse Representation Theory: Dynamic semantics with discourse structure. In Harry Bunt & Reinhard Muskens (eds.), *Computing Meaning, Volume 3*, 87–124. Kluwer.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177. doi:10.1016/ j.cognition.2007.05.006.

Mak, Willem M. & Ted J. M. Sanders. 2010. Incremental discourse processing: How coherence relations influence the resolution of pronouns. In Martin Everaert, Tom Lentz, Hannah De Mulder, Oystein Nilsen & Arjen Zondervan (eds.), *The Linguistics Enterprise*, 167–182. Benjamins.

Mak, Willem M. & Ted J. M. Sanders. 2013. The role of causality in discourse processing: Effects of expectation and coherence relations. *Language and Cognitive Processes* 28(9). 1414–1437. doi:10.1080/01690965.2012.708423.

Mann, William & Sandra Thompson. 1986. *Rhetorical Structure Theory: Description and Construction of Text Structures*. Information Sciences Institute.

Marchal, Marian, Merel Scholman, Ted Sanders & Vera Demberg. 2024a. What processing instructions do connectives provide? Modeling the facilitative effect of the connective. *Annual Meeting of the Cognitive Science Society* 46. 3435–3441.

Marchal, Marian, Merel C. J. Scholman & Vera Demberg. 2024b. How statistical correlations influence discourse-level processing: Clause type as a cue for discourse relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 50(5). 796–807. doi:10.1037/xlm0001270.

McHugh, Dean. 2023. *Causation and Modality*: University of Amsterdam PhD dissertation.

Mozuraitis, Mindaugas & Daphna Heller. 2017. Discourse coherence and the interpretation of accented pronouns. *Dialogue and Discourse* 8. 84–104. doi:10.5087/dad.2017.204.

Mulder, Gerben. 2008. *Understanding Causal Coherence Relations*: Utrecht University PhD dissertation.

Murray, John D. 1997. Connectives and narrative text: The role of continuity. *Memory & Cognition* 25(2). 227–236. doi:10.3758/bf03201114.

Nouwen, Rick, Adrian Brasoveanu, Jan van Eijck & Albert Visser. 2022. Dynamic semantics. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University Fall 2022 edn.

Pickering, Martin J., Brian McElree, Steven Frisson, Lillian Chen & Matthew J. Traxler. 2006. Underspecification and aspectual coercion. *Discourse Processes* 42(2). 131–155. doi:10.1207/s15326950dp4202_3.

Polanyi, Livia. 1985. A theory of discourse structure and discourse coherence. *Chicago Linguistic Society* 21. 306–322.

Pritchett, Bradley L. 1988. Garden Path phenomena and the grammatical basis of language processing. *Language* 64(3). 539–576. doi:10.2307/414532.

Pyykkönen, Pirita & Juhani Järvikivi. 2010. Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology* 57(1). 5–16. doi:10.1027/1618-3169/a000002.

Rohde, Hannah, Roger Levy & Andrew Kehler. 2011. Anticipating explanations in relative clause processing. *Cognition* 118(3). 339–358. doi:10.1016/j.cognition.2010.10.016.

Sasaki, Kelsey. 2021. *Components of Coherence*: UC Santa Cruz PhD dissertation.

Sasaki, Kelsey & Daniel Altshuler. 2025. Coherence relations: Inferences at the discourse, sentence, and clause-internal levels. Under review for *Language & Linguistics Compass*.

Scholman, Merel C. J., Hannah Rohde & Vera Demberg. 2017. 'On the one hand' as a cue to anticipate upcoming discourse structure. *Journal of Memory and Language* 97. 47–60. doi:10.1016/j.jml.2017.07.010.

Smyth, Ron. 1994. Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research* 23(3). 197–229. doi:10.1007/BF02139085.

Solstad, Torgrim & Oliver Bott. 2017. Causality and causal reasoning in natural language. In Michael R. Waldmann (ed.), *The Oxford handbook of causal reasoning*, Oxford University Press. doi:10.1093/oxfordhb/9780199399550.013.32.

Steedman, Mark & Gerry T. Altmann. 1989. "Ambiguity in context": A reply. *Language and Cognitive Processes* 4(3-4). SI105–SI122. doi:10.1080/01690968908406365.

Stevenson, Rosemary, Alistair Knott, Jon Oberlander & Sharon McDonald. 2010. Interpreting pronouns and connectives: Interactions among focusing, thematic roles, and coherence relations. *Language and Cognitive Processes* 15. 225–262. doi:10.1080/016909600386048.

Stewart, Andrew J., Martin J. Pickering & Anthony J. Sanford. 2000. The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language* 42. 423–443. doi:10.1006/jmla.1999.2691.

Sturt, Patrick & Matthew W. Crocker. 1997. Thematic monotonicity. *Journal of Psycholinguistic Research* 26(3). 297–322. doi:10.1023/A:1025076608204.

Traxler, Matthew J., Anthony J. Sanford, Joy P. Aked & Linda M. Moxey. 1997. Processing causal and diagnostic statements in discourse. *Journal of Ex-*

perimental Psychology: Learning, Memory, and Cognition 23(1). 88–101. doi:10.1037/0278-7393.23.1.88.

Webber, Bonnie. 2004. D-LTAG: Extending lexicalized TAG to discourse. *Cognitive Science* 28(5). 751–779. doi:10.1207/s15516709cog2805_6.

Webber, Bonnie, Matthew Stone, Aravind Joshi & Alistair Knott. 2003. Anaphora and discourse structure. *Computational Linguistics* 29. 545–587. doi:10.1162/089120103322753347.

Wolf, Florian, Edward Gibson & Timothy Desmet. 2004. Discourse coherence and pronoun resolution. *Language and Cognitive Processes* 19(6). 665–675. doi:10.1080/01690960444000034.

Xiang, Ming & Gina Kuperberg. 2015. Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience* 30(6). 648–672. doi:10.1080/23273798.2014.995679.

Xu, Xiaodong, Xiaoming Jiang & Xiaolin Zhou. 2015. When a causal assumption is not satisfied by reality: Differential brain responses to concessive and causal relations during sentence comprehension. *Language, Cognition and Neuroscience* 30(6). 704–715. doi:10.1080/23273798.2015.1005636.

Zeevat, Henk. 2011. Rhetorical relations. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An International Handbook of Natural Language Meaning*, 946–68. Mouton de Gruyter.

Zufferey, Sandrine & Liesbeth Degand. 2024. *Connectives and Discourse Relations*. Cambridge University Press. doi:10.1017/9781108966573.

John Duff
UCLA
Department of Linguistics
335 Portola Plaza
3125 Campbell Hall
Los Angeles, CA 90095
duff@ucla.edu

Daniel Altshuler
University of Oxford
Linguistics, Philology & Phonetics
Schwarzman Centre for the Humanities
Radcliffe Observatory Quarter
Woodstock Road
Oxford OX2 6GG
daniel.altshuler@ling-phil.ox.ac.uk

Kelsey Sasaki
Pomona College
Linguistics & Cognitive Science
333 N. College Way
Claremont, CA 91711
kelsey.sasaki@pomona.edu