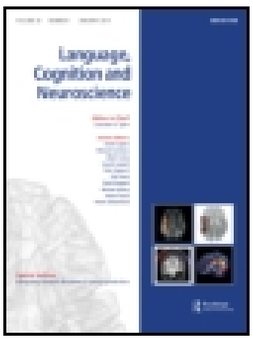

Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Scolding the child who threw the scissors: Shaping discourse expectations by restricting referents

Jet Hoek , Hannah Rohde , Jacqueline Evers-Vermeul & Ted J. M. Sanders

To cite this article: Jet Hoek , Hannah Rohde , Jacqueline Evers-Vermeul & Ted J. M. Sanders (2020): Scolding the child who threw the scissors: Shaping discourse expectations by restricting referents, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2020.1852292

To link to this article: https://doi.org/10.1080/23273798.2020.1852292

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2020.1852292

Routledge
Taylor & Francis Group

REGULAR ARTICLE

OPEN ACCESS Check for updates

Scolding the child who threw the scissors: Shaping discourse expectations by restricting referents

Jet Hoek[a], Hannah Rohde[b], Jacqueline Evers-Vermeul[a] and Ted J. M. Sanders[a]

[a]Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, Netherlands; [b]The University of Edinburgh, Edinburgh, UK

ABSTRACT
Coherence relations are often assumed to hold between clauses, but restrictive relative clauses (RCs) are usually not granted discourse segment status because they are syntactically and conceptually integrated in their matrix clauses. This paper investigates whether coherence relations can be inferred between restrictive RCs and their matrix clauses. Three experiments provide converging evidence that restrictive RCs can indeed play a role at the discourse level and should not categorically be excluded from receiving discourse segment status in discourse annotation practices. At the same time, the studies provide new insights into implicit causality verb biases, specifically about next-mention biases in concessive coherence relations, and expectations about discourse structure, upcoming referents, and upcoming coherence relations.

KEYWORDS
Coherence relations; relative clauses; discourse segments; discourse coherence; implicit causality

1. Introduction

Establishing meaning in a discourse depends on the inference of multiple types of dependencies. Context-driven processing can be said to require integration of material that has been encountered in the preceding context in order to make predictions about where the discourse will go next and which referents will be mentioned next. This interplay between coherence and coreference underpins a variety of studies on discourse structure and discourse processing (e.g. Asher & Lascarides, 2003; Hobbs, 1979; Kehler, 2002; Kehler et al., 2008; Koornneef & Sanders, 2013; Mak & Sanders, 2013). However, such work typically studies the way that clauses make their independent contribution to the establishment of a coherent discourse, focussing on one-clause sentences, main clauses, and subordinate clauses headed by a conjunction (see (1a-b) for examples of causal relations with the cause underlined). In the current work, we focus on smaller segments, restrictive relative clauses (RCs) as in (1c), whose role has often been overlooked but which may nonetheless provide content that contributes to the discourse structure and influences comprehenders' inferences about coherence and coreference.

(1a) I scolded the boy. *He stole a pencil case*.
(1b) I scolded the boy because *he stole a pencil case*.
(1c) I scolded the boy *who stole the pencil case*.

1.1. Discourse coherence

During discourse processing, language users keep track of many different types of information. When presented with new linguistic input, comprehenders integrate the new information into their representation of the discourse; anaphoric expressions are resolved to given discourse entities (e.g. *he* in (1a-b) easily resolving to *the boy*) and coherence relations are inferred between a new proposition and the preceding discourse. A coherence relation is an aspect of meaning that extends beyond the meaning of the clauses or sentences in isolation (Sanders et al., 1992), and an inferred coherence relation becomes part of the discourse representation that serves as a point of departure for processing upcoming linguistic information.

Both theoretical and corpus-based research highlight different types of coherence relations which language users can infer (e.g. Asher & Lascarides, 2003; Carlson & Marcu, 2001; Hobbs, 1990; Kehler, 2002; Prasad et al., 2007; Reese et al., 2007; Sanders et al., 1992; Wolf & Gibson, 2005). Much less research has investigated which parts of a discourse language users infer coherence relations between (notable exceptions are Hoek et al., 2018; Matthiessen & Thompson, 1988; Polanyi, 1988; Schilperoord & Verhagen, 1998; Verhagen, 2001).

Existing experimental studies on coherence relations appear to be influenced by common definitions of discourse segments, whether they work within a specific

CONTACT Jet Hoek j.hoek@let.ru.nl
Supplemental data for this article can be accessed at https://doi.org/10.1080/23273798.2020.1852292.

theory of discourse structure (e.g. Canestrelli et al., 2013; Kamalski et al., 2008), use corpus-based data annotated within a specific famework as the basis for their experiments Scholman & Demberg, 2017), or do not seem to assume a specific framework at all (e.g. Köhne & Demberg, 2013; Xiang & Kuperberg, 2015). Our goal of experimentally testing theories and practices of discourse segmentation thus not only evaluates the psychological validity of discourse annotation practices, but also informs experimental work on discourse coherence. Given that a range of computational work is based on corpus annotation (e.g. Lin et al., 2014; Muller et al., 2012; Sporleder & Lascarides, 2007) and draws insights from psycholinguistic studies (e.g. Cardie, 2000), assessing the validity of assumptions about discourse segmentation is relevant for computational approaches to discourse and for automated systems that involve discourse-level phenomena.

1.2. Discourse segmentation and restrictive relative clauses

Most approaches to discourse annotation have taken the grammatical clause as the basis for identifying discourse segments (e.g. Evers-Vermeul, 2005; Mann & Thompson, 1988; Sanders & van Wijk, 1996; Wolf & Gibson, 2005). One type of clause, however, that is commonly excluded from receiving discourse segment status is the restrictive relative clause (e.g. Mann & Thompson, 1988; Reese et al., 2007; Sanders & van Wijk, 1996; Verhagen, 2001).

Restrictive RCs are syntactically linked to a noun and are part of the noun phrase (NP) itself. Unlike non-restrictive RCs, they provide crucial information about the noun they modify, without which the conceptualisation of the referent is incomplete. Clauses that contain a restrictive RC are therefore conceptually dependent on the RC and, as such, the RC, the noun it modifies, and the rest of the clause containing that NP are assumed to form an integrated whole instead of independent discourse segments (Schilperoord & Verhagen, 1998). While non-restrictive RCs, as in (2), are traditionally considered to be discourse segments, restrictive RCs, as in (3), thus seem to be excluded as discourse segments because their contribution is defined instead by their role in ensuring that the matrix clause's status is a referentially (and therefore propositionally) fully specified clause.

(2) Susan, who is brilliant, now works at NASA.
(3) Someone I knew in high school now works at NASA.

In some restrictive RC constructions, however, one can see that a coherence relation could plausibly be inferred between the restrictive RC and its matrix clause. Examples of such constructions are (4)–(6), as is also illustrated by the paraphrases in (4′)–(6′).

(4) Man who attacked jogger in Seattle park sentenced to prison.[1]
(4′) Man is sentenced to prison because he attacked a jogger in Seattle park.
(5) 'I. Did. Not. Plagiarize. That. Paper,' Laura Pottsdam says of the paper that was almost entirely plagiarised.[2]
(5′) 'I. Did. Not. Plagiarize. That. Paper,' Laura Pottsdam says of her paper, even though it was almost entirely plagiarised.
(6) Anyone who cares about food should be eating in Texas.[3]
(6′) if you care about food, you should be eating in Texas.

Another indication that it is indeed possible, at least sometimes, to infer a coherence relation between a restrictive RC and its matrix clause, comes from translation. The Europarl Direct corpus (Cartoni et al., 2013; Koehn, 2005), for instance, contains examples of coherence relations between freestanding clauses that were translated using a restrictive RC construction, as in (7), and vice versa, as in (8). In both (7) and (8), the overall meaning of the translation is similar to the meaning of the original English fragment. The clauses between which the coherence relation in (7) holds have been put in bold.

(7) EN Recently we have seen headlines in Dutch and Irish newspapers about *jet aircraft being chartered to fly workers from the west of Ireland to jobs in the Netherlands* because *the Netherlands cannot get workers to do this work*.
NL Onlangs meldden Nederlandse en Ierse kranten dat er *vliegtuigen werden gecharterd om arbeiders uit het westen van Ierland naar Nederland te vervoeren voor banen* waar *geen Nederlandse werknemers voor kunnen worden gevonden*.
'…*for jobs* for which *no Dutch employees could be found*.'
(8) EN However, those consular services are not available to Muslims from other EU MemberStates *who* would be there under the same terms and conditions operated by the Saudi authorities as UK Muslims.
DE Zu diesen Einrichtungen haben jedoch Muslime aus anderen EU-Mitgliedstaaten keinen Zugang, *obwohl* für diese dieselben Vorschriften der saudischen Behörden gelten, wie für die Muslime aus dem Vereinigten Königreich.

'…although *they would be there under the same terms* …'

Examples (4)–(8) appear to extend the types of clauses between which coherence relations can be inferred. In all examples, the restrictive RC seems to contribute to the discourse structure beyond specifying the referent of the noun it modifies. As such, they may influence comprehenders' expectations about upcoming coherence relations and about upcoming referents. How these RCs influence expectations about upcoming discourse will most likely not be uniform, but rather determined by the preceding context in combination with the coherence relation that is inferred between the RC and its matrix clause; the examples in (4)–(8) suggest that there is a range of coherence relations that restrictive RCs can enter into.

There is some preliminary evidence from psycholinguistic studies that suggest that comprehenders indeed infer coherence relations between restrictive RCs and their matrix clauses. However, prior studies have only considered causal relations, similar to (4) and (7), and have done so in ambiguous contexts that may have necessitated comprehenders to search for discourse-level information in clauses where they would not otherwise look for it. The current experiments test whether comprehenders also infer relations other than causal relations between restrictive RCs and their matrix clauses, whether the inference of coherence relations in restrictive RC constructions occurs even in contexts without ambiguity or specific task demands, and whether restrictive RCs can influence expectations about upcoming discourse.

2. Background: implicit causality and discourse expectations

When processing language, comprehenders are understood to generate predictions about upcoming discourse (for overviews of prediction in language processing, see Kuperberg & Jaeger, 2016; Kutas et al., 2011). At the level of discourse, language users have for instance been shown to have expectations about upcoming coreference (e.g. Arnold, 2001; Kehler et al., 2008; Koornneef & van Berkum, 2006), coherence relations (e.g. Rohde & Horton, 2014), and discourse structure, i.e. to which part of the preceding discourse a segment will attach (Scholman et al., 2017). Well-studied elements that have been shown to yield discourse-level expectations are Implicit Causality (IC) verbs. IC verbs are known to have both coherence and coreference biases, see Section 2.1, and will be used as a basis for the manipulations in our studies. The psycholinguistic studies that provide preliminary evidence for a role for restrictive RCs in discourse interpretation have also made use of IC verbs, though none of the studies were directly aimed at investigating restrictive RCs, see Section 2.2.

2.1. Coherence and coreference biases in IC contexts

Implicit causality verbs are transitive verbs that assign special status to the referent that comprehenders associate with the cause of the situation depicted (Au, 1986; Ferstl et al., 2011; Garvey & Caramazza, 1974; Kehler et al., 2008; Koornneef & Sanders, 2013; Koornneef & van Berkum, 2006; Mak & Sanders, 2013; McKoon et al., 1993), among many others). Story continuation tasks for prompts like (9)–(10) reveal that participants have strong preferences regarding which referent to mention next. The *next-mention biases* are specifically found in explanations, prototypically marked by *because* (e.g. Hartshorne, 2014; Kehler et al., 2008; Pickering & Majid, 2007; Solstad & Bott, 2013). NP1-biased IC verbs favour the subject, as in (11); NP2-biased IC verbs favour the object, as in (12).

(9) Tracy annoyed Tom (because) …
(10) Tracy fired Tom (because) …
(11) **Tracy** annoyed Tom because **she** kept complaining.
(12) Tracy fired **Tom** because **he** kept complaining.

A few studies have looked at coreference biases of IC verbs in other coherence relations, for instance consequence or contrast relations (Commandeur, 2010; Kehler et al., 2008; Koornneef & Sanders, 2013; Pickering & Majid, 2007; Stevenson et al., 2000). These studies find that an IC verb's coreference bias is influenced by the type of coherence relation it enters into with the following clause. However, all these studies have focussed on coherence relations between either free-standing clauses or clauses linked by a conjunction (*because*, *but*, *and*),[4] and an open question is whether IC verb coreference patterns are malleable in discourse contexts where a relation is inferred with the content of a restrictive RC. In addition, no study has investigated IC verb coreference patterns in negative causal relations (also called concessive relations or denial of expectation relations, prototypically signalled by *although* or *even though*), see (5′) and (8).

Another property of IC verbs crucial to the experiments reported in this paper is that they raise the expectation of an upcoming explanation. In a continuation task using prompts similar to (9)–(10) without *because*, Kehler et al. (2008) show that IC verbs receive

about 60% explanation continuations, while only 24% of continuations following non-IC verbs constitute an explanation.

There has been a lot of debate about what exactly gives rise to IC biases, but most accounts seem to conclude that the biases are the product of (pragmatic) inferences (Hartshorne, 2014; Kehler et al., 2008; Pickering & Majid, 2007, among others). Here we test if these discourse-level inferences are influenced by restrictive RCs that can be related to their matrix clauses beyond merely providing referential information about one of the arguments, and if this influence differs depending on the type of relation that can be inferred between the RC and its matrix clause.

2.2. Relative clauses in IC contexts

There is some evidence that suggests that restrictive RCs can enter into a coherence relation with their matrix clause. In a continuation experiment, Rohde et al. (2011) presented participants with prompts containing IC verbs and asked them to continue a relative clause that could be attached to two potential referents, as in (13); in this example, *the children* and *the musician* compete for RC modification.

(13) John detests the children of the musician who …

In the experiment, participants often used the RC to supply an explanation for the main clause verb, sometimes in a way that appeared to restrict the referent of the noun 'children' or 'musician.' However, since the contents of the RCs were supplied by the participants, restrictiveness cannot be guaranteed.

In a continuation experiment conducted by Kehler and Rohde (2019), participants were presented with prompts containing an IC verb and an RC. Participants were found to supply fewer explanations after an RC if a causal relation could be inferred between the RC and its matrix clause, as in (14), than after prompts where the RC merely provided additional information about its referent, as in (15).[5]

(14) The boss fired the employee who was embezzling money.
(15) The boss fired the employee who was hired in 2002.

Like Rohde et al.'s (2011) continuation task, this experiment was also not specifically focussed on restrictive RCs. While the RCs in (14) and (15) can plausibly be interpreted as restrictive, not all items make a restrictive reading possible or plausible. For instance, in an item like *The onlooker complimented the bride who…* it is unlikely that there are multiple potential bride-referents.

In addition to prior work not being specifically focussed on coherence relations with restrictive RCs, evidence is still lacking for the real-time inference of such relations in natural comprehension. Story continuation tasks require participants to actively engage with linguistic material for which they do not have any prior context. In having to build a coherent story out of a constrained scenario, participants may have drawn inferences beyond those they would have drawn in a less demanding comprehension task. Where there has been an investigation of online processing (Rohde et al., 2011), the items contained syntactic ambiguity as in (13), the resolution of which may have required participants to engage in additional inferencing. Moreover, these items were again not specifically aimed at testing restrictive RCs.

2.3. The current study

The studies discussed above seem to provide preliminary evidence supporting the idea that restrictive RCs can enter into a coherence relation with their matrix clauses. However, an unanswered question is whether language users' discourse-level inferences for restrictive RCs are limited to contexts that invoke specific task demands or require the resolution of syntactic ambiguity, or whether it is possible that inferring coherence relations between restrictive RCs and their matrix clauses is a process that occurs more generally. In addition, any effects found in prior experiments may have been driven by items that contained non-restrictive RCs. Finally, all prior experiments are limited to causal coherence relations between RCs and their matrix clauses. This paper addresses all three issues.

In all our experiments, we test restrictive RCs only. While the distinction between restrictive and non-restrictive RCs in not always entirely clear-cut (e.g. Bache & Jakobsen, 1980), we designed the RCs in all three experiments to be more characteristic of restrictive RCs than of non-restrictive RCs, using criteria listed in Bache and Jakobsen (1980) and Fabb (1990), among others: the RC is not separated from its matrix clause by means of a comma, the matrix clause and the RC can plausibly be uttered as a single intonation unit, the relative pronoun *who* can plausibly be substituted with *that*, and the RC cannot be removed from the sentence without losing essential information. In addition, the RC modifies the noun so that it refers to a unique referent (Experiment 1 and 2) or picks out a unique referent from a mentioned or invoked set of possible referents (Experiment 3).

Experiments 1 and 2 test whether a restrictive RC can enter into a relation with its matrix clause beyond causal (explanation) relations, as is suggested by examples (5), (6), (8). Since conditional interpretations, such as the one in (6), seem most plausible in contexts that contain a quantifier, we instead focus on the availability of negative causal relations (also called denial of expectation relations); for ease of reference, these relations will be referred to as 'concessive' relations. Experiment 1 and 2 (like the story continuation studies reviewed in Section 2.2) use off-line measures to explore the possibility of coherence relations between restrictive RCs and their matrix clauses. Experiment 3 then uses a self-paced reading paradigm to test whether restrictive RC constructions show processing behaviour similar to that of other types of coherence relations between independent clauses.

3. Continuation experiment 1

Experiment 1 tests whether restrictive RCs can influence next-mention expectations of the subsequent sentence. It aims to replicate the finding by Kehler and Rohde (2019) that RCs that provide a plausible cause for the matrix clause event can influence next-mention expectations, and to investigate whether restrictive RCs that provide an implausible cause (i.e. concessive RCs) can do the same. In this study, we presented participants with prompts for which they were asked to supply a natural continuation. Target prompts consisted of a matrix clause containing an NP2-biased IC verb, an object modified by an RC, and a connective. Prompts differed in the coherence relation that could be inferred between the RC and the matrix clause (causal, concessive, or neutral), and in the connective (*because* or *even though*), see (16)–(18).

(16) We thanked the neighbour [who brought over a fruit basket]$_{causalRC}$
  a. *because*
  b. *even though*
(17) We thanked the neighbour [who stopped by on Tuesday night]$_{neutralRC}$
  a. *because*
  b. *even though*
(18) We thanked the neighbour [who dropped our newly inherited vase]$_{concessiveRC}$
  a. *because*
  b. *even though*

In (16)–(18), each NP2-biased IC verb construction includes a restrictive RC modifying the object. If restrictive RCs are indeed available for discourse-level inferences, continuations should be malleable given the content of the RC. On the other hand, if restrictive RCs only contribute to the meaning of the sentence by restricting reference, any possible links between the RC and the matrix clause should be irrelevant to subsequent next-mention biases. In other words, if we find a main effect of RC condition, or potentially a connective × RC interaction, this would be in line with the idea that restrictive RCs can have a function at the discourse level. on participants' choice of next mention this would be in line with the idea that restrictive RCs can have a function at the discourse level. More specifically, we predict that if a causal relation is inferred between the restrictive RC and the main clause, as in (16), the IC bias (i.e. an explanation featuring the NP2) should be fulfilled (Kehler & Rohde, 2019). The next-mention bias following (16) is thus predicted to differ from (17) and (18). Below we describe the possible patterns for *because* and *even though*.

For *because* prompts, the NP2 bias is expected to be reduced in the causal RC condition (16a) compared to the neutral RC condition in (17a). The prediction is a bit more complex for the concessive RC condition (18). A concessive relation between the restrictive RC and its matrix clause, as in (18), indicates that something unexpected happens; thanking someone for ruining an heirloom is not a standard event. This discrepancy warrants an explanation. Compared to a neutral NP2-biased IC verb construction, as in (17a), there are multiple relevant candidates to focus an explanation on in (18a); the explanation may focus on the NP2 (e.g. they offered to replace it), but also on the NP1 (e.g. we are too nice for our own good), or on some other factor (e.g. the vase may have been incredibly ugly). Concessive RCs may thus reduce the NP2 bias for subsequent clauses although not necessarily to the same extent as causal RCs. The main reason why predictions about next-mention biases after a concessive RC are less straightforward than predictions for the causal condition is that much less is known about next-mention expectations after an IC verb in a concessive context than after IC verbs that feature in a causal relation. Although several studies have explored IC biases after *but* (e.g. Ehrlich, 1980; Koornneef & Sanders, 2013), we have not found any papers that specifically deal with IC verbs in concessive relations. We included the *even though* prompts to investigate the effect that a concessive context has on next-mention biases. The neutral+*even though* condition (17b) serves as a baseline. We predict the NP2 bias in this condition to be reduced as compared to the neutral+*because* condition (17a). *Even though* signals that there is something unexpected going on, which, in the context of an NP2-biased IC

verb may boost the relevance of the NP1, since that referent is doing something unusual. Because IC biases do not pertain to concession, we expect the RC type manipulation to have less of an effect in the *even though* condition (16b), (17b), (18b) than in the *because* condition (16a), (17a), (18a).

3.1. Participants

56 monolingual English speakers were recruited through Amazon Mechanical Turk (mean age 34.88, age range 23–66, 20 women). They participated in exchange for monetary compensation ($6.00).

3.2. Materials

Participants were presented with 30 target prompts consisting of a matrix clause containing an NP2-biased IC verb, an object that was modified by an RC, and a connective, see (16)–(18). The IC verbs used in all experiments reported in this paper were taken from existing inventories of IC verbs (Commandeur, 2010; Ferstl et al., 2011; Koornneef & van Berkum, 2006). The subject of all stimuli was a proper name (80%) or a first person pronoun (20%); the direct object was a definite NP whose gender was specified (e.g. guy) or relied on stereotypical gender assignments (e.g. a gardener is usually male) and which differed from the gender of the subject referent. For items with proper name subjects, the subject was male in 50% of the items and female in the other 50%. The NP2-biased IC verb always appeared in the past tense. The full list of target items can be found in Appendix 1.

The target prompts were distributed over six lists, with each item occurring only once per list, in one of the six conditions. The 30 target prompts were interspersed with 16 fillers containing a connective, an embedded clause, or both, and 24 fillers from an unrelated experiment. The items from each list were presented to the participants in random order.

3.3. Procedure

Continuations were collected via a web-based interface embedded in the Amazon Mechanical Turk environment. Each item was displayed on a separate page. Participants were instructed to write a natural continuation for the prompts in the supplied text box. Beforehand, the participants were informed that the experiment would not take longer than an hour; on average, participants took approximately 45 minutes to complete the experiment.

3.4. Annotation and data preparation

One trained coder (first author) annotated all continuations for the referent of the subject of the continuation: NP1, as in (19a), NP2, as in (19b), or Other, as in (19c).

(19) Natalie distrusted the doctor who had messed up the procedure last time because
  a. **she** could have died.
  b. **he** didn't seem to own up to his previous mistake.
  c. **such a breach of trust** was hard to shake.

Being a relatively simple task, this type of annotation is generally reliable, especially in contexts where the referents have different (specified or stereotyped) genders. We double-coded (author and another trained annotator) the continuations for Experiment 2. The annotation of Experiment 2 indeed shows a very high agreement for next-mention at 96%, $\kappa = .94$ (see also Section 4.3).

The annotation process revealed a subset of unfinished continuations and continuations that were completely nonsensical (4%). For the analysis, we only included continuations in which the connective attached to the main clause (87%) because these allow us to assess the effect of the RC on subsequent material that attaches to the main event being described in the mini-discourse.

3.5. Analysis method

All experiments in this paper were analysed using linear mixed effects regression models (LMER: Baayen, 2008; Baayen et al., 2008) or, in case of categorical dependent variables, generalised linear mixed effects regression models (GLMM), using the lme4 package (Bates et al., 2015) in R (R Development Core Team, 2016). Models contained fixed effects for RC type, connective type, and their interaction, as well as by-participant and by-item random effects. For each model, we started with a maximal random effects structure, only simplifying the model in case of nonconvergence (Barr et al., 2013). We first reduced the random effects by taking out correlations between (either the by-participant or by-item) random slopes and random intercepts. If the model still did not converge, we iteratively removed random slopes until we ended up with a converging model. See Barr et al. (2013) for a detailed account of this step-wise procedure.

The significance of fixed effects was determined by performing likelihood ratio tests to compare the fit of the model to that of a model with the same random

effects structure that did not include the fixed effect. The categorical predictor variables in all analyses were deviation coded. All pairwise comparisons were obtained using a subset of the data that only contained the relevant conditions with re-centred predictor variables. For example, to assess the significance of the three-level factor RC type, we compare models with and without the RC type factor. If the model with RC type is found to be significantly better, we conduct follow-up analyses assessing pairs of RC types (e.g. causal vs. concessive) to identify the source of the main effect of RC type.

3.6. Results

The proportions of NP1, NP2, and Other continuations per condition are shown in Figure 1. In our analysis, we modelled the binary outcome of NP2 versus non-NP2 continuations in a generalised mixed effects model. In keeping with a model in which participants infer coherence relations between matrix clauses and restrictive RCs and such inferences in turn influence coreference, we found a main effect of RC type ($p < .001$). In addition, we find a main effect of connective, whereby, as predicted, there were fewer NP2 continuations after *even though* than after *because* ($\beta = -0.52$, $SE = 0.19$, $z = -2.87$, $p < .01$). Even though Figure 1 suggests a potential dampening of the effect of RC type in the *even though* condition, the interaction between RC type and connective was not significant ($p = .22$).

Pairwise comparisons for RC type reveal that, in line with our hypotheses, there were fewer NP2 continuations after a causal RC than after a concessive RC ($\beta = -1.13$, $SE = 0.47$, $z = -2.43$, $p < .05$) or a neutral RC ($\beta = -1.09$, $SE = 0.22$, $z = -4.94$, $p < .001$). There were also fewer NP2 continuations after a concessive RC than after a neutral RC ($\beta = -0.51$, $SE = 0.20$, $z = -2.55$, $p < .05$).

3.7. Discussion

The results from the continuation study show that behaviour varies with the content of the RC, in keeping with an account in which coherence relations can hold between RCs and their matrix clauses. Causal RCs lead to the strongest reduction of the next-mention bias, but concessive RCs also reduce the proportion of references to the NP2.

Similar to concessive RCs, a concessive connective reduces the next-mention bias of IC verbs. The reduction in NP2 bias in the *even though* condition affected all RC conditions (there was no interaction effect between RC type and connective). This indicates that the influence of the RC and the connective on the next-mention bias is cumulative; a concessive RC reduces the expectation of the NP2 being mentioned as the subject of the subsequent clause, after which the concessive connective further reduces this NP2 bias.

In sum, the continuation study shows distinct next-mention patterns after causal, concessive, and neutral RCs. This suggests that language users indeed infer coherence relations between restrictive RCs and their matrix clauses.

4. Continuation experiment 2

Experiment 1 tested whether restrictive RCs can influence expectations about upcoming referents. When coding the continuations, we noticed that not all continuations attached to the main clause; some participants linked their continuation to the contents of the RC. The difference between these two construction types is illustrated by comparing (20), where *because* most plausibly attaches to the main clause, and (21), where *because* can be understood to begin an explanation for the contents of the RC. As such, (20) and (21) have distinct discourse structures, namely [[CLAUSE1$_{MATRIX}$ + CLAUSE2$_{RC}$] *because* [CLAUSE3$_{EXPLANATION}$]] and [CLAUSE1$_{MATRIX}$ [[CLAUSE2] *because* [CLAUSE3$_{EXPLANATION}$]]$_{RC}$]. We will refer to attachments to the main clause as high attachments, and to attachments within the RC as low attachments.

(20) *Mrs. Thompson loathed the gardener* who never took off his muddy shoes *because he made a mess all throughout her beautiful home.*

(21) Mrs. Thompson loathed the gardener *who never took off his muddy shoes because he was embarrassed of his foot odour.*

Experiment 2 asks whether restrictive RCs can guide expectations about discourse structure. As before, we hypothesise that if a causal relation is inferred between the RC and its matrix clause, there would no longer be an expectation for upcoming causal information to explain the matrix-clause IC verb event. We would then expect any further causal cues to favour attachment to another part of the discourse, for instance the RC, compared to when the IC causal requirement has not yet been fulfilled. In other words, we expect fewer high attachments of *because* after a causal RC than after a neutral RC.

Since our discourse structure predictions apply specifically to explanation contexts, the prompts

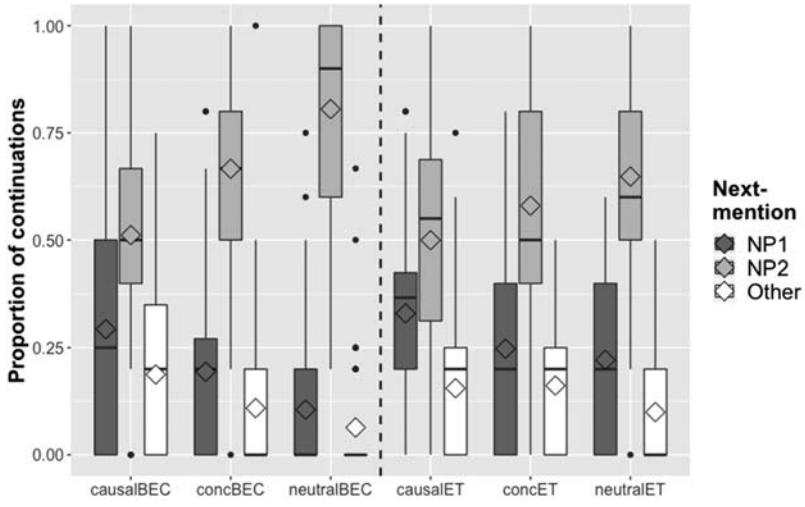

Figure 1. Proportion of next-mentioned referents per connective (left side: BECause; right side: Even Though) and RC type. The bold horizontal lines indicate the medians; the tilted squares indicate the means.

used in this experiment only include *because* as a connective. Regarding RC types (causal, neutral, concessive), we do not necessarily expect concessive RCs to impact the discourse structure differently than the neutral RCs, but we kept the concessive RC condition to check if the next-mention results from Experiment 1 are replicable.

4.1. Participants

55 monolingual English speakers were recruited through Amazon Mechanical Turk (mean age 38.25, age range 22–67, 26 women). They participated in exchange for monetary compensation ($5.50).

4.2. Materials

The 30 target items in Experiment 2 were the same as the items in Experiment 1, with the exception of the connective manipulation; only the *because* versions were included in this experiment. 24 fillers were created to replace the fillers from the unrelated experiment in Experiment 1. The new fillers all contained an embedded structure. Half of the items required or were biased toward high attachment, i.e. an attachment of the continuation to the matrix clause; for instance *so* in (22a) most prototypically signals a result/consequence relation, which is usually not embedded (Asher & Vieu, 2005; Hoek et al., 2017). The other half required or were biased toward low attachment, i.e. a continuation within the embedded clause, as in (22b); continuations for this prompt should provide the second segment of the embedded conditional relation. The attachment manipulation in the fillers was intended to prevent an overall bias toward high or low attachments. In Experiment 1, all of the fillers had been biased toward high attachment.[6]

(22a) Wade insisted that penguins did not really exist so …
(22b) The manual stated that if the red light in the top right corner of the dryer was blinking …

The target prompts were distributed over three lists, with each item occurring only once per list, in one of the three conditions. The 30 target prompts were interspersed with the 24 attachment fillers and 16 additional fillers of various types. The items from each list were presented to the participants in random order. The procedure followed that of Experiment 1.

4.3. Annotation and data preparation

Two trained coders (first author and an undergraduate Linguistics student) annotated all continuations for the referent of the subject of the continuation, using the categories NP1, NP2, and Other (see also Section 3.4). The agreement between the coders was very high: 96%, $\kappa = .94$.

In addition, we annotated for each continuation whether it attached to the matrix clause, as in (23a) or to the RC, as in (23b). We also included a label 'both' for continuations that could as plausibly be attached to the matrix clause as to the RC, or to both at the same time, as in (23c). In cases where the content of the continuation appears to relate to both the RC and the matrix clause, the continuation seems to syntactically attach within the RC, but conceptually also relates

to the matrix clause. Continuations classified as 'both' attachments seem to primarily be examples of causal chaining, rather than instances of truly ambiguous attachment.

(23) Geoff ridiculed the stewardess who crashed the drink cart into the wall because …
  a. he was that kind of person.
  b. she was dizzy during the flight.
  c. she was so clumsy.

Annotating the attachment of a continuation is a more complex task than annotating co-reference in contexts with referents of different genders, since determining attachment relies more heavily on interpretation. Agreement between the two coders was satisfactory at 94% and $\kappa = .74$. Disagreements were resolved through discussion.

We removed any unfinished continuations, as well as continuations that were completely nonsensical (1%). For our next-mention analysis, we included only high attachments (89%); low attachment continuations attach to the RC and, as such, the relation marked by *because* does not contain the IC verb in its relational segments; this makes continuations about the subject of the matrix clause (NP1) unavailable. In attachments coded as 'both,' the IC verb is included in one of the two relations marked by *because*, but since this discourse structure is distinctly different from the discourse structure found in high attachments, we excluded such cases from our next-mention analysis to keep our dataset as homogeneous as possible. For our attachment analysis, we included all finished/sensible continuations.

4.4. Results

The next-mention results replicate the results from Experiment 1, as shown in Figure 2. We used generalised linear mixed effects regression to model the binary outcome of NP2 versus not-NP2 continuations and found a main effect of RC type ($p < .001$). Pairwise comparisons again revealed that there were fewer NP2 continuations after causal RCs than after concessive RCs ($\beta = -1.03$, $SE = 0.27$, $z = -3.18$, $p < .001$) and neutral RCs ($\beta = -1.94$, $SE = 0.32$, $z = -6.15$, $p < .001$). Again, there were fewer NP2 continuations after concessive RCs than after neutral RCs ($\beta = -0.90$, $SE = 0.28$, $z = -2.27$, $p < .01$).

Figure 3 shows the proportion of high, low, and 'both' attachments. We analysed the binary outcome of high versus not-high attachments using generalised linear mixed effects regression modelling. We found a main effect of RC type ($p < .001$). Pairwise comparisons revealed that, as predicted, there were fewer high attachments to the matrix clause after causal RCs than after both concessive RCs ($\beta = -2.30$, $SE = 0.59$, $z = -3.92$, $p < .001$) and neutral RCs ($\beta = -3.38$, $SE = 0.82$, $z = -4.10$, $p < .001$). There was no difference between concessive RCs and neutral RCs ($\beta = -1.46$, $SE = 1.26$, $z = -1.16$, $p = .25$).

4.5. Discussion

The next-mention results in Experiment 2 replicated the next-mention results from Experiment 1. The attachment results are in line with the prediction for fewer high attachments in the causal condition than in the other two conditions. However, we expected this effect to be mainly driven by an increase in the number of low attachments in the causal condition. Even though the results do indicate that participants provided more low attachments in the causal condition than in the concessive or the neutral condition, it was more the increase in 'Both' continuations (those that related to both the matrix clause and the RC) that reduced the number of high attachments in the causal condition, as can be seen in Figure 3. Overall, however, the distinct attachment patterns for the different conditions support our more global hypothesis that language users infer coherence relations between restrictive RCs and their matrix clauses.

5. Experiment 3: self-paced reading

Experiment 1 and 2 investigated the influence of restrictive RCs on expectations about the continuation of the discourse. The results of both studies corroborate the observation that readers can infer causal coherence relations between restrictive RCs and their matrix clauses, and established that other types of coherence relations are available as well. However, both experiments use off-line measures and so provide no information about the time course with which the coherence relations are inferred. Are these results a reflection of some aspect of the story continuation task, or do readers draw these inferences naturally during reading? This question is addressed in Experiment 3. By means of a self-paced reading task, we investigate whether the processing of coherence relations that hold between restrictive RCs and their matrix clauses mirrors the processing of coherence relations in more traditional constructions, e.g. between two independent clauses. The self-paced reading experiment in Rohde et al. (2011) already provides evidence that language users can expect RCs to convey causal

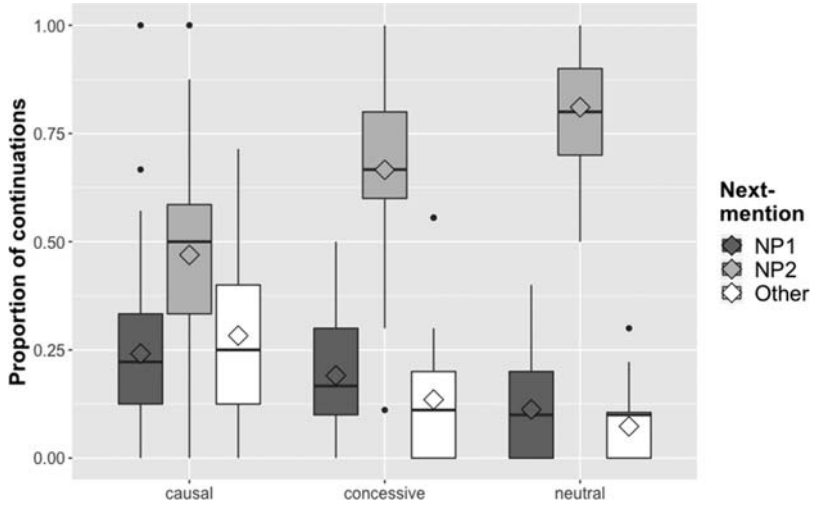

Figure 2. Proportion of next-mentioned referents per RC type. The bold horizontal lines indicate the medians; the tilted squares indicate the means.

information. The current experiment further investigates the availability of causal inferences, now specifically in restrictive RCs. In addition, it also tests the inference of concessive relations between restrictive RCs and their matrix clauses. In contrast to prior work, our items do not involve the disambiguation of the relative pronoun; as such, this experiment examines whether the Rohde et al. (2011) findings were mainly due to participants using cues from the discourse to help disambiguate the referent of the relative pronoun, or whether restrictive RCs are generally places where language users expect discourse-level information.

A well-established finding in discourse processing is that causal information is processed faster than non-causal information, and that stronger causal links result in even faster processing times than weaker causal links (e.g. Haberlandt & Bingham, 1978; Keenan et al., 1984; Myers et al., 1987; Sanders & Noordman, 2000; Wolfe et al., 2005). Indeed, several studies report slower reading times on relations where there is some form of contrast between the discourse segments (e.g. adversative, concessive, or contrastive relations) than relations where there is not (e.g. Koornneef & Sanders, 2013; Lee & Lee, 2005; Wason & Johnson-Laird, 1972). These findings also seem to hold true in the context of IC verbs: Causal relations after an IC verb lead to faster reading times than additive or negative relations (Koornneef & Sanders, 2013; Mak & Sanders, 2013).

If coherence relations between restrictive RCs and their matrix clauses are processed in a way that mirrors the processing of coherence relations between independent clauses, causal RCs should be read faster than neutral RCs, which should in turn be read faster than concessive RCs. If, however, restrictive RCs represent

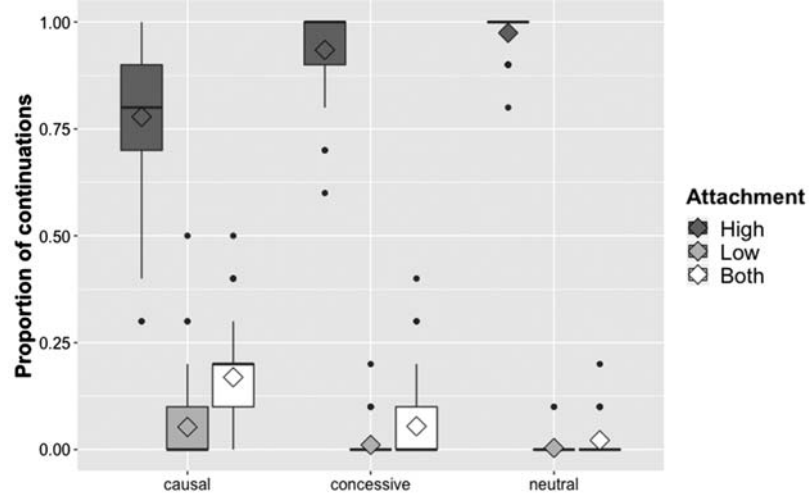

Figure 3. Proportion of attachments per RC type. The bold horizontal lines indicate the medians; the tilted squares indicate the means.

linguistic constructions in which language users do not typically expect to find information that is relevant at the discourse level, reading times should be fastest for neutral RCs since those RCs can be understood as simply disambiguating the referent; in contrast, reading times would be slowed by RCs whose content makes available a coherence relation and this pragmatic enrichment of meaning is posited to take time (akin to the delay reported for the computation of implicatures, e.g. Huang & Snedeker, 2009).

We thus expect concessive RCs to be most surprising and most difficult. It should, however, be noted that although concessive RCs occur in everyday language use, see (5), it is bound to be much rarer than causal and neutral RCs. Concessive relations are more complex and less expected by language users than causal or additive relations (e.g. Hoek et al., 2017). In addition, they are much less often expressed without an overt linguistic marker (Asr & Demberg, 2012; Hoek et al., 2017; Taboada, 2006). Fine et al. (2013, p. 2) formulate the 'rapid expectation adaptation' account, which states that 'comprehenders are able to rapidly adapt to the statistics of novel linguistic environments.' In two self-paced reading tasks, they show that the processing disadvantage of linguistic constructions that are usually rare diminishes or even disappears if participants are repeatedly exposed to such constructions in an experimental setting. Applied to our experiment, this account predicts that concessive RCs will be read slowly in the beginning of the experiment, but that reading times for this condition speed up as the experiment progresses. To test this, we include a main effect of trial number in our models and test whether the interaction between trial number and RC type is significant.

It has also been found that the content of causal relations is verified faster and recalled better than information from clauses that are not part of a causal relation (e.g. Sanders & Noordman, 2000; Trabasso & van den Broek, 1985; van den Broek, 1990). In addition to comparing reading times of causal, neutral, and concessive RCs, Experiment 3 measures whether information provided by causal RCs is verified faster than information provided by neutral or concessive RCs. The verification statements only inquire about the contents of individual clauses, not about any discourse-level inferences, to avoid influencing participants' reading behaviour. Since the resulting statements are fairly easy to verify and participants who are paying attention ought to be at ceiling, we only analyze reaction times, not the accuracy of responses.

We do use the accuracy of responses to check whether participants were paying attention during the experiment, requiring a minimum accuracy of 75% on the verification statements. None of the participants scored below this threshold. In general, we assume that measuring reading times can be done reliably using web-based experiments, as is for instance shown by studies comparing findings from web-based and in-lab reaction time experiments (e.g. Enochson & Culbertson, 2015; Keller et al., 2009).

5.1. Participants

52 monolingual English speakers were recruited through Amazon Mechanical Turk (mean age 40.22, age range 25–63, 31 women). They participated in exchange for monetary compensation ($4.50).

5.2. Materials

Stimuli contained an introductory sentence that introduced or invoked a set of people from which one person would later be singled out by the restrictive RC construction, a target sentence consisting of a matrix clause with a direct object modified by an RC, and a wrap-up sentence. The target sentences varied in the coherence relation that could be inferred between the RC and the matrix clause (causal, neutral or concessive). The subject of all stimuli was a proper name or a first person pronoun, while the direct object was a general NP that specified or implied a different gender or person than the subject; for subjects with proper names, the subject was male in 50% of the items, female in the other 50%. The verb in the matrix clause was always an NP2-biased IC verb in the past tense. Each IC verb was matched with another IC verb to create (context-dependent) antonyms, e.g. *admire* and *pity*, or *thank* and *sue*. By manipulating the IC verb to change the coherence relation between the matrix clause and the RC, the RC was kept constant between conditions, see Table 1. Each set of IC verb antonyms was supplemented with a non-IC verb to create a neutral condition with the same RC. Each IC verb was used twice: in one item it occurred in the causal condition, in another item in the concessive condition.

Table 1. Sample item with target sentence in all three conditions.

Intro	Jenny walked through the hallway to check on the daily goings-on around the office.
neutral RC	She **joked with** the guy who made a lot of money for the company.
causal RC	She **praised** the guy who made a lot of money for the company.
concessive RC	She **fired** the guy who made a lot of money for the company.
Wrap-up	She arrived at the conference room just in time for her next meeting.
Verification statement	*The guy made a lot of money for the company.*

Each participant saw each verb only once. The full list of target items can be found in Appendix 2.

The target items were distributed over three lists, with each item occurring only once per list, in one of the three conditions. 24 target items were interspersed with 12 'distractor' fillers that also contained RCs but which were not systematically manipulated for the type of relation that could be inferred between the RC and the matrix clause,[7] and 24 additional fillers of various types.[8] Each participant saw every item only once, in one of the conditions.

Each item was accompanied by a verification statement. For the target items, the verification statement inquired only about the content of the RC and was always true. For the 'distractor' fillers that also contained RCs, the statements also always asked about the contents of the RC but were always false. For the additional fillers, the statements were a mix of true and false, and asked about various parts of the stimuli. In total, a third of all verification statements were false; two thirds were true (see also Appendix 2).

5.3. Procedure

Participants were recruited via Amazon Mechanical Turk, after which they were directed to another website, hosted by IbexFarm (Drummond, 2013), where they completed the moving window self-paced reading experiment. Items were initially displayed as a series of horizontal lines on the screen; the length of the lines corresponded to the length of the regions. By pressing the space bar on their keyboard, participants could reveal the next region of the item. Items were presented non-cumulatively; when a new region was revealed, the previous region was again replaced by lines.

All target sentences were split up into two regions, with the matrix clause and the relative pronoun in the first region, and the rest of the sentence (RC content) in the second region. The first and last sentences of every item were also presented as two regions. Each target sentence started on a new line and was followed by the first region of the wrap-up sentence. (24) illustrates the spatial configuration of target stimuli on the screen, with slashes demarcating regions.

(24) Jenny walked through the hallway to check on / the daily goings-on around the office
She praised the guy who / made a lot of money for the company. / She arrived at the conference room / just in time for her next meeting.

When finished reading the item, participants had to press the space bar once more to move on to the verification statement. They responded to the statement by clicking either TRUE or FALSE with their cursor. After six randomly selected items, participants were presented with a picture of a landscape. These pictures allowed participants to take a short break without it affecting the reading time measures. When they were ready to continue, participants clicked a 'proceed' button at the bottom of the screen.

5.4. Data preparation and analysis

For the analysis, we used residual reading times. Residual reading times were calculated using a regression equation that predicts the reading time of a region based on a participant's reading speed and the length of the region; the predicted reading time is then subtracted from the actual reading time of the region (Trueswell et al., 1994). Residual reading times thus adjust for differences in the length of regions as well as differences in participants' reading rates. Negative residual reading times indicate that a region was read faster than predicted, positive residual reading times indicate that a region was read slower than predicted. We removed residual reading times that were more than three standard deviations above or below the mean (0.52% of the data). RTs at three regions were analysed: pre-target (matrix), target (RC), and spillover (wrap-up sentence). These were analysed with linear mixed effect models with fixed effects for RC type, trial number, and their interaction. For significance testing, we use likelihood ratio tests that compare two models that are identical except for the inclusion/exclusion of one fixed effect.

All participants' verification statement accuracy was above chance. As expected, the average percentage of correct responses was very high (93.94%), with 96.15% accurate responses to the target items and 92.47% to the filler items. The accuracy of responses per subject ranged between 76.67% and 100%. The reading time analysis was performed on all non-outlier data, regardless of whether the participant answered the item's verification statement correctly. The analysis of the reaction time to the verification statements was performed on correct responses only.

5.5. Results

5.5.1. Reading times

Table 2 provides an overview of the raw reading times per condition on the matrix clause, the RC, and the first region of the wrap-up sentence, which we have labelled as the spill-over region. Figure 4 shows the residual reading times on each of these regions for all three conditions.

Table 2. Mean raw reading times and standard deviations per condition per region in milliseconds.

	Matrix clause		*RC*		*Spill-over*	
	M	SD	M	SD	M	SD
neutral RC	1243	667	1573	903	1302	763
causal RC	1207	695	1500	918	1271	702
concessive RC	1189	547	1706	1149	1350	813

We analysed the residual reading times on the RC region in a linear mixed effects model. We found significant main effects of trial number ($p < .001$) and RC type ($p < .05$), as well as a significant interaction between trial number and RC type ($p < .05$). This indicates that while reading times sped up over the course of the experiment, the progression of the experiment impacted the three conditions differently. To interpret this effect, we divided the dataset into the first and second half of the experiment, and plotted the residual reading times for both halves, see Figure 5. Throughout the experiment, reading times were fastest on the causal RCs. It seems that concessive RCs were initially read slower than both causal and neutral RCs, but as the experiment progressed, the difference in reading time between neutral and concessive RCs diminished, which can also been seen from the raw reading times for the RC region between the first and second half of the experiment in Table 3. Indeed, separate follow-up analyses of the first and the second half of the experiment indicate a significant main effect of the three-level condition variable in both halves (first half: $p < .001$, second half: $p < .01$). While causal RCs were read faster than the other two RC types in both halves of the experiment (all *ps* $< .01$), reading times between the neutral and concessive condition (obtained through pairwise analyses on the relevant subsets of the data) only differ in the first half of the experiment ($\beta = 129.52$, $SE = 45.43$, $t = 2.85$, $p < .01$), but not in the second half ($\beta = 17.58$, $SE = 41.74$, $t = 0.42$, $p = .67$).

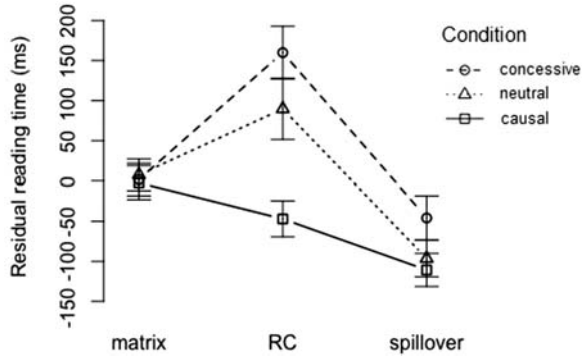

Figure 4. Residual reading times on the matrix clause, the RC, and the spill-over region per condition, in milliseconds.

On both the matrix clause and the spill-over region, there was a main effect of trial number ($p < .01$) There was also a main effect of RC type on the spill-over region ($p < .05$). Pairwise comparisons on the relevant subsets of the data reveal that the spill-over region after concessive RCs was read slower than the spill-over region after causal RCs ($\beta = -65.63$, $SE = 29.94$, $t = -2.20$, $p < .05$). There was no difference between the causal and the neutral ($\beta = -12.46$, $SE = 25.97$, $t = -0.48$, $p = .63$) or between the neutral and the concessive condition ($\beta = 52.97$, $SE = 29.09$, $t = 1.82$, $p = .08$). There was no main effect of RC type on the matrix clause ($p = .95$). Interactions between trial number and RC type were also not significant on the matrix clause and spill-over regions (matrix: $p = .66$, spill-over: $p = .76$).

5.5.2. Reaction times on verification statements

Only accurate responses were included in the analysis of the reaction times of the verification statements (96.15% of the data; 97.12% for causal RCs, 95.67% for concessive and neutral RCs). Table 4 shows the mean reaction times to the verification statements per condition. There was no main effect of condition ($p = .68$).

5.6. Discussion

The results of the self-paced reading experiment are in line with findings regarding the relative processing

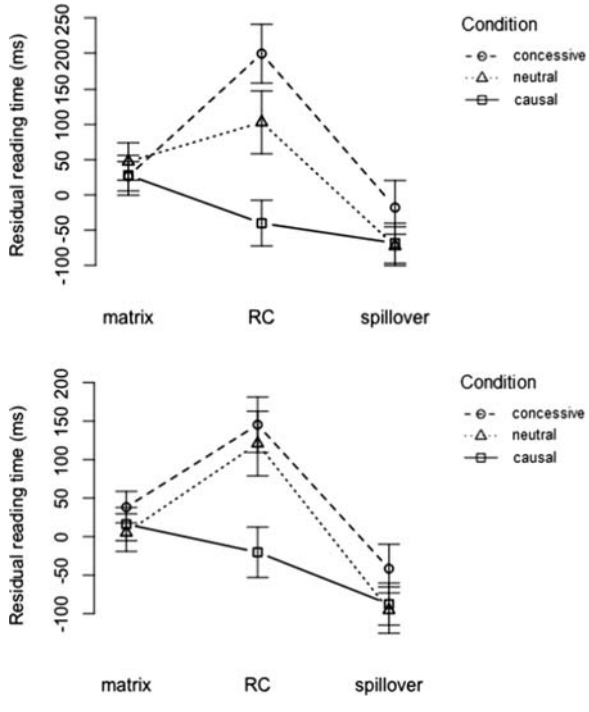

Figure 5. Residual reading times on the matrix clause, the RC, and the spill-over region per condition in milliseconds in the first (above) and second half (below) of the experiment.

Table 3. Mean raw reading times and standard deviations per condition for the RC region for the first and the second half of the experiment, in milliseconds.

	First half		*Second half*	
	M	SD	M	SD
neutral RC	1589	861	1514	846
causal RC	1486	789	1361	694
concessive RC	1744	892	1515	796

ease of causal, additive, and concessive relations that hold between independent clauses. In addition, we found an interaction between trial number and condition on the RC region. The reading time plots in Figure 5 and raw reading times in Table 3 show that the biggest speed-up effect occurred in the concessive condition. As explained above, concessive relations are more complex and less expected by language users than causal or additive relations (e.g. Hoek et al., 2017) and are much less often expressed without an overt linguistic marker (Asr & Demberg, 2012; Hoek et al., 2017; Taboada, 2006). With the biggest speed-up effect found in the rarest of our three conditions, it seems that the rapid expectation adaptation account by Fine et al. (2013) forms a plausible explanation for the interaction effect we found.

Overall, the reading time results are in line with processing studies on coherence relations between independent clauses, rather than with reading time patterns that would be predicted if language users do not expect restrictive RCs to contain information that is relevant at the discourse level. This suggests that readers naturally infer coherence relations between restrictive RCs and their matrix clauses, and that they do not only start making these inferences when the linguistic context (i.e. syntactic disambiguation) or the experimental setting encourages them to do so (i.e. story continuations).

We found no differences between the conditions in the speed with which the statements were verified. This may be due to the way in which participants had to respond to the statements. After they finished reading the items, they pressed the spacebar one more time to move to the verification statement. They then had to respond to the statements by clicking on either the TRUE or FALSE button with their cursor. It was not possible in the web-hosted experiment to control the position of the participants' cursor as they entered the verification screen. The cursor may have been anywhere, and plausibly at the edge of the screen so that it would not cover part of the self-paced reading items (participants were instructed to move their cursors out of the way). This, in combination with the time it took participants to switch from the keyboard to the mouse, probably resulted in a fairly noisy measurement and may have made it more difficult to detect any differences between conditions.

Table 4. Mean reaction times and standard deviations of verification statements per condition, in milliseconds.

	M	SD
neutral RC	2202	946
causal RC	2205	923
concessive RC	2250	998

6. General discussion and conclusion

Over the course of three experiments, this paper explored whether coherence relations can be inferred to hold between restrictive RCs and their matrix clauses. The experiments provide converging evidence that suggests that language users treat restrictive RCs as linguistic elements that potentially contain information that is relevant at the discourse level, and that the contents of a restrictive RC can be linked to its matrix clause in a causal or concessive, i.e. negative causal, coherence relation. This implies that restrictive RCs should not be categorically excluded from receiving discourse segment status. At the same time, the studies provided new insights into implicit causality verb biases, specifically about next-mention biases in concessive coherence relations, and expectations about discourse structure, upcoming referents, and upcoming coherence relations.

The assumption that restrictive RCs cannot enter into a coherence relation with their matrix clauses or any other clauses was based on the fact that restrictive RCs are both syntactically and conceptually integrated in their matrix clauses. The RC is syntactically embedded in an NP and, as such, part of the matrix clause. In addition, the RC provides vital information about the referent it modifies, without which the matrix clause is conceptually incomplete. The experiments in this paper suggest that neither of these factors make the restrictive RC completely unavailable as a discourse segment. This finding suggests that there are limitations to syntax-based criteria for discourse segmentation; underlying the idea that restrictive RCs cannot function as discourse segments because it is syntactically embedded in its matrix clause is the assumption that discourse structure and syntactic structure align. However, in the causal and concessive experimental items, as well as in examples (2)–(6) given in the introduction, there appears to be a mismatch between the syntactic structure and the discourse structure of the sentence. In (25), for example, the RC relates to only the NP *the guy*

at the syntactic level, while at the discourse level it relates to the entire matrix clause.

(25) She praised the guy who made a lot of money for the company.

When it comes to the conceptual level, the RC appears to fulfil a double function. It restricts the referent so that, in the case of (25), it is clear which guy is praised. At the same time, it provides a reason or explanation for the matrix clause event, in this case the praising. The observation that syntactic structure and discourse structure need not align and that a single linguistic unit can fulfil multiple functions open up the larger question of whether we should be looking for discourse-level information in other types of linguistic structures that have been assumed to be unavailable at the discourse level on the basis of syntactic or conceptual criteria.

If restrictive RCs cannot be categorically excluded as discourse segments, how should they be treated in discourse segmentation and annotation? A liberal option would be to allow every restrictive RC to be a discourse segment. In the absence of a meaningful discourse-level relation between the RC and its matrix clause, the relation could be annotated as ADDITIVE (or using a similarly general relation label, such as ELABORATION (RST/Hobbs), CONJUNCTION (PDTB), or BACKGROUND (RST/SDRT)). The drawback of this approach is that the discourse structure and discourse annotations may include relations that are irrelevant at the discourse level; while the experiments in this paper show that a restrictive RC *can* relate to its entire matrix clause at the discourse level, there is no reason to believe that this is *always* the case. A more conservative option would be to only segment a restrictive RC construction and annotate the coherence relation between the segments if the sentence allows for a discourse-level inference between the RC and its matrix clause. Since this segmentation procedure relies heavily on the interpretation of the annotator, it is not a very suitable option for automated segmentation or annotation applications if these systems are blind to the relation to be inferred. For manual segmentation and annotation procedures, however, this option would likely result in a more accurate representation of the discourse structure and a more complete overview of the coherence relations that hold in a discourse.

In this study, we focussed on restrictive RCs because their status as grammatical clauses and full propositions would grant them discourse segment status, if only they related to another clause at the level of syntax. This makes them a border-line case within theories on discourse segmentation. However, it could be questioned to what extent a construction like *I scolded the boy who stole a pencil case* differs from a sentence like *I scolded the thief*, as both constructions would receive similar causal inferences.[9] *I scolded the thief*, however, would be straightforwardly analysed as a single discourse segment, and the causal inference between the scolding and the thief-being would not be considered a discourse relation.The reason why theories on discourse segmentation take the grammatical clause as the basis for identifying discourse segment is because the clause is used as a proxy for identifying independent propositions in a text (Hoek et al., 2018). Perhaps theories on discourse segmentation should start to consider the role of *inferred* propositions. This would for instance problematise the status of *thief* in the above example (similarly: nominalizations, e.g. *arrest*), but also question distinct analyses of constructions such as *The blindfolded man could not find his chair* (one discourse segment) versus *The man could not find his chair (because) he was blind-folded* (two discourse segments).

Throughout all three experiments, we tested RCs that were attached to a direct object. The surface structure of these constructions closely resembles the surface structure of coherence relations with more prototypical discourse segments: clause-clause. An open question is whether coherence relations are equally likely to be inferred between restrictive RCs and their matrix clauses when the RC interrupts the matrix clause, for instance when the RC is attached to the subject. In addition, all RCs in our experiments were subject-extracted RCs, where the referent of the RC is the subject of the RC. Object-extracted RCs, where the referent of the RC is the object of the RC (e.g. *the stewardess who Geoff ridiculed*), are syntactically more complex and tend to be harder to process than subject-extracted RCs (Gibson, 1998; Gordon et al., 2001; Mak et al., 2002). Whether discourse-level inferences are made as readily in constructions with object-extracted RCs as in sentences containing subject-extracted RCs would be worthwhile to explore in future research.

Notes

1. https://www.seattletimes.com/seattle-news/man-who-attacked-jogger-in-seattle-park-sentenced-to-prison/
2. Hill, N. (2016). The Nix. London: Picador. p.36
3. https://www.eater.com/2018/3/7/17080432/texas-dining-barbecue-mexican-crawfish
4. Studies by Kehler et al. (2008) and Mak and Sanders (2013) do show that IC verbs' coreference patterns are influenced by the inferred coherence relation, not just by the conjunction.
5. This study was first presented in brief in Kehler and Rohde (2015).

6. Interestingly, this manipulation seems not to have impacted the proportion of high attaching continuations at all (89% in Experiment 1 versus 87% in Experiment 2). This suggests that people attend more to the matrix clause (see also Larson, 2008).
7. For example: *The painter was busy making the final arrangements for her exposition. She refused to talk to the journalist who was trying to interview her. After two hours he finally gave up and left.* Verification statement: *The journalist tried to photograph the painter.*
8. For example: *Sunday afternoon was pretty eventful. After our bunny escaped, we spent two hours looking for him. Finally, we found him hiding under the couch, sitting in a spot that was impossible to reach.* Verification statement: *The bunny was hiding under a chair.*
9. Thanks to the anonymous reviewer for pointing this out.

Acknowledgments

This research was supported by the SNSF Sinergia project MODERN (CRSII2_147653) and by a Leverhulme Trust Prize in Languages and Literatures to H. Rohde.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the SNSF (Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung) Sinergia project MODERN (CRSII2_147653) and by a Leverhulme Trust Prize in Languages and Literatures to H. Rohde (PLP-2016-128).

References

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, *31*(2), 137–162. https://doi.org/10.1207/S15326950DP310202

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.

Asher, N., & Vieu, L. (2005). Subordinating and coordinating discourse relations. *Lingua*, *115*, 591–610. https://doi.org/10.1016/j.lingua.2003.09.017

Asr, F. T., & Demberg, V. (2012). Implicitness of discourse relations. In M. Kay & C. Boitet (eds.), *Proceedings of the 24th International Conference on Computational Linguistics (COLING), Mumbai, India* (pp. 2669–2684). The COLING 2012 Organizing Committee.

Au, T. K.-F. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language*, *25*(1), 104–122. https://doi.org/10.1016/0749-596X(86)90024-0

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bache, C., & Jakobsen, L. K. (1980). On the distinction between restrictive and non-restrictive relative clauses in modern English. *Lingua*, *52*(3–4), 243–267. https://doi.org/10.1016/0024-3841(80)90036-4

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–67. https://doi.org/10.18637/jss.v067.i01

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

Cardie, C. (2000). A cognitive bias approach to feature selection and weighting for case-based learners. *Machine Learning*, *41*(1), 85–116. https://doi.org/10.1023/A:1007665204628

Carlson, L., & Marcu, D. (2001). Discourse tagging reference manual. *ISI Technical Report ISI-TR-545*, *54*, 1–56.

Cartoni, B., Zufferey, S., & Meyer, T. (2013). Annotating the meaning of discourse connectives by looking at their translation: The translation-spotting technique. *Dialogue & Discourse*, *3*(2), 65–86. https://doi.org/10.5087/dad.2013.204

Commandeur, E. (2010). *Implicit causality and implicit consequentiality in language comprehension*. Number 14 in TiCC Dissertation Series. Gildeprint Drukkerijen, Enschede, The Netherlands.

Drummond, A. (2013). Ibex farm. Online server: http://spellout.net/ibexfarm

Ehrlich, K. (1980). Comprehension of pronouns. *The Quarterly Journal of Experimental Psychology*, *32*(2), 247–255. https://doi.org/10.1080/14640748008401161

Enochson, K., & Culbertson, J. (2015). Collecting psycholinguistic response time data using amazon mechanical turk. *PloS One*, *10*(3), e0116946. https://doi.org/10.1371/journal.pone.0116946

Evers-Vermeul, J. (2005). *The development of Dutch connectives: Change and acquisition as windows on form-function relations*. LOT.

Fabb, N. (1990). The difference between English restrictive and nonrestrictive relative clauses. *Journal of Linguistics*, *26*(1), 57–77. https://doi.org/10.1017/S0022226700014420

Ferstl, E. C., Garnham, A., & Manouilidou, C. (2011). Implicit causality bias in English: A corpus of 300 verbs. *Behavior Research Methods*, *43*(1), 124–135. https://doi.org/10.3758/s13428-010-0023-2

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS One*, *8*(10), 1–18. https://doi.org/10.1371/journal.pone.0077661

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

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *27*(6), 1411–1423. https://doi.org/10.1037/0278-7393.27.6.1411

Haberlandt, K., & Bingham, G. (1978). Verbs contribute to the coherence of brief narratives: Reading related and unrelated sentence triples. *Journal of Verbal Learning and Verbal Behavior*, *17*(4), 419–425. https://doi.org/10.1016/S0022-5371(78)90247-5

Hartshorne, J. K. (2014). What is implicit causality? *Language, Cognition and Neuroscience*, *29*(7), 804–824. https://doi.org/10.1080/01690965.2013.796396

Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, *3*(1), 67–90. https://doi.org/10.1207/s15516709cog03014

Hobbs, J. R. (1990). *Literature and cognition*. CSLI Publications.

Hoek, J., Evers-Vermeul, J., & Sanders, T. J. M. (2018). Segmenting discourse: Incorporating interpretation into segmentation? *Corpus Linguistics and Linguistic Theory*, *14*(2), 357–386. https://doi.org/10.1515/cllt-2016-0042

Hoek, J., Zufferey, S., Evers-Vermeul, J., & Sanders, T. J. M. (2017). Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study. *Journal of Pragmatics*, *121*, 113–131. https://doi.org/10.1016/j.pragma.2017.10.010

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415. https://doi.org/10.1016/j.cogpsych.2008.09.001

Kamalski, J., Sanders, T. J. M., & Lentz, L. (2008). Coherence marking, prior knowledge, and comprehension of informative and persuasive texts: Sorting things out. *Discourse Processes*, *45*(4–5), 323–345. https://doi.org/10.1080/01638530802145486

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 115–126. https://doi.org/10.1016/S0022-5371(84)90082-3

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI Publications.

Kehler, A., Kertz, L., Rohde, H., & J. L. Elman (2008). Coherence and coreference revisited. *Journal of Semantics*, *25*(1), 1–44. https://doi.org/10.1093/jos/ffm018

Kehler, A., & Rohde, H. (2015). Pronominal reference and pragmatic enrichment: A Bayesian account. In R. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society (CogSci2015), Pasadena, CA, USA* (pp. 1063–1068). Cognitive Science Society.

Kehler, A., & Rohde, H. (2019). Prominence and coherence in a Bayesian theory of pronoun interpretation. *Journal of Pragmatics*, *154*, 63–78. https://doi.org/10.1016/j.pragma.2018.04.006

Keller, F., Gunasekharan, S., Mayo, N., & Corley, M. (2009). Timing accuracy of web experiments: A case study using the webexp software package. *Behavior Research Methods*, *41*(1), 1–12. https://doi.org/10.3758/BRM.41.1.12

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *Proceedings of the Tenth Machine Translation Summit (MT Summit X), Phuket, Thailand* (pp. 79–86). Asia-Pacific Association for Machine Translation.

Köhne, J., & Demberg, V. (2013). The time-course of processing discourse connectives. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society (CogSci), Berlin, Germany*. Cognitive Science Society.

Koornneef, A. W., & Sanders, T. J. M. (2013). Establishing coherence relations in discourse: the influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, *28*(8), 1169–1206. https://doi.org/10.1080/01690965.2012.699076

Koornneef, A. W., & van Berkum, J. J. (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*(4), 445–465. https://doi.org/10.1016/j.jml.2005.12.003

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford University Press.

Larson, M. (2008). Structural priming from complex sentences. In *Proceedings from the Annual Meeting of the Chicago Linguistics Society 44.1, Chicago, USA* (pp. 179–193). Chicago Linguistic Society.

Lee, J.-M., & Lee, J.-H. (2005). Contrastive information processing in discourse comprehension. *Korean Journal of Cognitive Science*, *16*(2), 1–24.

Lin, Z., Ng, H. T., & Kan, M.-Y. (2014). A pdtb-styled end-to-end discourse parser. *Natural Language Engineering*, *20*(2), 151–184. https://doi.org/10.1017/S1351324912000307

Mak, W. M., & T. 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. https://doi.org/10.1080/01690965.2012.708423

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, *47*(1), 50–68. https://doi.org/10.1006/jmla.2001.2837

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, *8*(3), 243–281. https://doi.org/10.1515/text.1.1988.8.3.243

Matthiessen, C., & Thompson, S. A. (1988). The structure of discourse and 'subordination'. In J. Haiman & S. Thompson (Eds.), *Clause combining in grammar and discourse* (pp. 275–329). John Benjamins.

McKoon, G., Greene, S. B., & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1040–1052. https://doi.org/10.1037/0278-7393.19.5.1040

Muller, P., Afantenos, S., Denis, P., & Asher, N. (2012). Constrained decoding for text-level discourse parsing. In M. Kay & C. Boitet (eds.), *Proceedings of the 24th International Conference on Computational Linguistics (COLING), Mumbai, India* (pp. 1883–1900). The COLING 2012 Organizing Committee.

Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, *26*(4), 453–465. https://doi.org/10.1016/0749-596X(87)90101-X

Pickering, M. J., & Majid, A. (2007). What are implicit causality and consequentiality? *Language and Cognitive Processes*, *22*(5), 780–788. https://doi.org/10.1080/01690960601119876

Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics*, *12*(5–6), 601–638. https://doi.org/10.1016/0378-2166(88)90050-1

Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A. K., Robaldo, L., & Webber, B. (2007). *The Penn Discourse Treebank 2.0 annotation manual*.

R Development Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Reese, B., Hunter, J., Asher, N., Denis, P., & Baldridge, J. (2007). *Reference Manual for the Analysis and Annotation of Rhetorical Structure (Version 1.0)*.

Rohde, H., & Horton, W. S. (2014). Anticipatory looks reveal expectations about discourse relations. *Cognition*, *133*(3), 667–691. https://doi.org/10.1016/j.cognition.2014.08.012

Rohde, H., Levy, R., & Kehler, A. (2011). Anticipating explanations in relative clause processing. *Cognition*, *118*(3), 339–358. https://doi.org/10.1016/j.cognition.2010.10.016

Sanders, T. J. M., & Noordman, L. G. M. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, *29*(1), 37–60. https://doi.org/10.1207/S15326950dp29013

Sanders, T. J. M., Spooren, W. P. M. S., & Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, *15*(1), 1–35. https://doi.org/10.1080/01638539209544800

Sanders, T. J. M., & van Wijk, C. (1996). PISA – a procedure for analyzing the structure of explanatory texts. *Text: Interdisciplinary Journal for the Study of Discourse*, *16*(1), 91–132. https://doi.org/10.1515/text.1.1996.16.1.91

Schilperoord, J., & Verhagen, A. (1998). Conceptual dependency and the clausal structure of discourse. In J. Koenig (Ed.), *Discourse and cognition: Bridging the gap* (pp. 141–163). CSLI Publications.

Scholman, M. C. J., & Demberg, V. (2017). Examples and specifications that prove a point: Identifying elaborative and argumentative discourse relations. *Dialogue & Discourse*, *8*(2), 56–83. https://doi.org/10.5087/dad.2017.203

Scholman, M. C. J., Rohde, H., & Demberg, V. (2017). "On the one hand" as a cue to anticipate upcoming discourse structure. *Journal of Memory and Language*, *97*, 47–60. https://doi.org/10.1016/j.jml.2017.07.010

Solstad, T., & Bott, O. (2013). Towards a formal theory of explanatory biases in discourse. In M. Aloni, M. Franke, & F. Roelofsen, *Proceedings of the 19th Amsterdam colloquium, Amsterdam, The Netherlands* (pp. 203–210). ILLC, University of Amsterdam.

Sporleder, C., & Lascarides, A.. (2007). *Exploiting linguistic cues to classify rhetorical relations*. In N. Nicolov, K. Bontcheva, G. Angelova, & R. Mitkov (eds.), Recent advances in natural language Processing IV: Selected papers from RANLP 2005 (pp. 157-166). John Benjamins.

Stevenson, R., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes*, *15*(3), 225–262. https://doi.org/10.1080/016909600386048

Taboada, M. (2006). Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics*, *38*(4), 567–592. https://doi.org/10.1016/j.pragma.2005.09.010

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*(5), 612–630. https://doi.org/10.1016/0749-596X(85)90049-X

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*(3), 285–318. https://doi.org/10.1006/jmla.1994.1014

van den Broek, P. (1990). Causal inferences and the comprehension of narrative texts. In A. Graesser & G. Bower (Ed.), *Psychology of learning and motivation: Inferences and text comprehension* (Vol. 25, pp. 175–196). Academic Press.

Verhagen, A. (2001). Subordination and discourse segmentation revisited, or: Why matrix clauses may be more dependent than complements. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psychological aspects* (pp. 337–357). John Benjamins.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Harvard University Press.

Wolf, F., & Gibson, E. (2005). Representing discourse coherence: A corpus-based study. *Computational Linguistics*, *31*(2), 249–287. https://doi.org/10.1162/0891201054223977

Wolfe, M. B., Magliano, J. P., & Larsen, B. (2005). Causal and semantic relatedness in discourse understanding and representation. *Discourse Processes*, *39*(2–3), 165–187. https://doi.org/10.1207/s15326950dp3902&34

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