

Conscious rereading is confirmatory: Evidence from bidirectional self-paced reading

Dario Paape, Department Linguistik, Universität Potsdam, Germany, paape@uni-potsdam.de

Shravan Vasishth, Department Linguistik, Universität Potsdam, Germany, vasishth@uni-potsdam.de

Rereading during sentence processing can be confirmatory, in which case it serves to increase readers' certainty in their current interpretation, or it can be revisionary, in which case it serves to correct a misinterpretation (Christianson et al., 2017). The distinction is particularly relevant in garden-path sentences, which have been argued to trigger revisionary rereading (Frazier and Rayner, 1982). In two web-based experiments that compare garden-path sentences with other linguistic constructions, we investigate deliberate rereading in the recently-proposed bidirectional self-paced reading (BSPR) paradigm (Paape and Vasishth, 2022). Our results show evidence for selective rereading only in very difficult garden-path sentences. Additionally, our results suggest that conscious, selective rereading is confirmatory: Readers find garden-path sentences less rather than more acceptable after selective rereading, suggesting that they reread either to confirm their initial analysis or to confirm the perceived ungrammaticality of the sentence. We discuss the role of conscious awareness in dealing with different types of linguistic inconsistency.



1. Introduction

During reading, 10% to 15% of readers' eye movements go against the forward flow of the text, that is, they are regressive (Rayner, 1998; Rayner et al., 2006). Regressions occur less frequently when previously read text is masked, suggesting that their purpose is to allow rereading (Booth and Weger, 2013; Schotter et al., 2014). The need for rereading can result from a failure to integrate the current word into the sentential context when it is highly unexpected or incongruous (e.g., Yan and Jaeger, 2020; Staub, 2011). When readers encounter an unexpected word (*Since Jay always jogs a mile and a half seems ...*), this may lower their confidence in the current interpretation of the previous sentence context (Bicknell and Levy, 2011; Weiss, 2020; Levy, 2008; Levy et al., 2009), and they may launch a regression to either confirm or revise their current interpretation (e.g., Christianson et al., 2017; Frazier and Rayner, 1982). If the reader's task is to judge a given sentence's syntactic and/or semantic well-formedness, rereading may serve to confirm a first impression (Paape et al., 2021; Godfroid et al., 2015a), resulting in higher judgment accuracy compared to trials without rereading (Metzner et al., 2017).

A topic that has received comparatively little attention in psycholinguistic studies of regressions and rereading is the role of conscious awareness. Skilled reading is largely considered to be a highly automatic process (Reichle et al., 1998; Logan, 1997), and conscious awareness of processing difficulty, which may result in deliberate rereading, has mostly been studied in groups whose processing is presumably less automatic, that is, in non-native speakers (see Godfroid et al., 2015b for a review) and in children (e.g., Eilers et al., 2018).

For instance, in an eye-tracking study, Hessel et al. (2021) investigated how 9- to 10-year-olds respond to inconsistent discourse continuations such as *kitten* in (1).

- (1) Rover barks at all passing animals on the street. He's the most alert kitten in the whole neighborhood, in fact.

Hessel et al. found that compared to a consistent discourse continuation (*puppy*), children were more likely to reread the word *kitten*, and also spent more time rereading it. Hessel et al. also found a correlation between self-reports of deliberate rereading strategies ("I go back to words and sentences that I have found difficult to understand when I first read them") and observed rereading of the *kitten/puppy* region in the eye-tracking record, but no interaction with discourse consistency. Hessel et al. concluded that deliberate rereading may play only a minor role in dealing with such inconsistencies, which may often be resolved in a largely automatic fashion (Van den Broek and Helder, 2017).

An open question is whether evidence for deliberate rereading can be observed in adult native speakers, with a different type of linguistic manipulation that does not target discourse consistency, and/or by using different tasks and reading paradigms. The distinction between *confirmatory* and *revisionary* rereading (Christianson et al., 2017) is of particular interest in this

context. In the sentences investigated by Hessel et al. (2021), children presumably engaged in confirmatory rereading, that is, they wanted to make sure that the sentence was indeed about a barking kitten. Revisionary rereading, on the other hand, has been argued to play a role in garden-path sentences such as (2):

(2) Since Jay always jogs a mile and a half seems like a short distance to him.

Here, the noun phrase *a mile and a half* is likely to be initially misanalyzed as the direct object of the verb *jogs* as opposed to the subject of a new clause. Once the verb *seems* is read, the sentence is disambiguated towards an analysis in which *jogs* is used intransitively. This type of syntactic ambiguity is known as the NP/zero complement (NP/Z) ambiguity. The *selective reanalysis hypothesis* proposed by Frazier and Rayner (1982) states that when the disambiguating information arrives, readers will preferentially reread those parts of the sentence that they misanalyzed on the first pass. This type of rereading is revisionary in the sense that it coincides with syntactic reanalysis. For (2), the prediction is that readers will return to the verb *jogs* and/or to the problematic noun phrase *a mile and a half* in order to find an alternative syntactic analysis, namely one in which *a mile and a half* is the subject of *seems* rather than the object of *jogs*.

Frazier and Rayner assume that this overt, selective process of syntactic reanalysis is the parser's default recovery strategy after being garden-pathed, and occurs in the absence of conscious awareness. They argue that because conscious awareness of garden-pathing is rare, it cannot reflect the parser's "fundamental response to the pervasive temporary ambiguity of natural languages" (p. 179), but rather take awareness as an index of failure to selectively reanalyze the sentence. Frazier and Rayner found some evidence of selective reanalysis in their eye-tracking study, but the evidence from later studies investigating selective reanalysis has been much more mixed (Mitchell et al., 2008; Meseguer et al., 2002; von der Malsburg and Vasishth, 2011, 2013; Schotter et al., 2014; Christianson et al., 2017, in prep.). Importantly, Christianson et al. (2017) and Christianson et al. (in prep.) found no relationship between rereading of garden-path sentences and comprehension accuracy: Participants often misinterpret a sentence like (2) to mean that *Jay always jogs a mile and a half* even after rereading it. This suggests that the purpose of rereading in garden-path sentences may be confirmatory rather than revisionary, similar to the *puppy/kitten* sentences of Hessel et al. (2021).

As a possible explanation for the mixed results in the literature, Paape and Vasishth (2022) suggested that selective reanalysis may be tied to conscious awareness of the garden path, contra Frazier and Rayner (1982). Depending on the syntactic construction, garden paths may be mild or strong, and may or may not rise to awareness. The NP/Z garden path in particular is relatively strong and may cause readers to consciously experience parsing failure. In such cases, it is possible that selective, truly revisionary rereading is observed, indexing a deliberate attempt to find a grammatical structure. Such deliberate rereading can be triggered by naturally-occurring

newspaper headlines such as *Google's computer might better translation tool* (Novick et al., 2014), which often cause the reader to do a double take and reread the sentence. Subjectively, in such difficult sentences, a switch occurs from automatic processing to a more deliberate, metacognitive process of reanalysis in which the comprehender actively tries to figure out where they went wrong (Marcus 1980; Lewis 1998; Gibson 1991).¹

A variety of scientific tools exist to study deliberate reading strategies, including self-reports and “thinking aloud” paradigms, which have the advantage of tracking strategy use as it occurs (Cromley and Azevedo, 2006). As an additional tool that requires neither introspection nor verbalization on part of participants, Paape and Vasishth (2022) proposed the bidirectional self-paced reading (BSPR) paradigm. BSPR is an extension of “moving-window” self-paced reading (SPR; Mitchell and Green, 1978; Aaronson and Scarborough, 1976), which does not allow regressions. By contrast, in BSPR, participants can move forward or backward through the sentence by pressing specific keyboard keys, for instance the right and left arrow keys. By pressing one of two additional keys, they can also return directly to the beginning of the sentence, or proceed directly to the end-of-trial task from any point in the sentence.² Paape and Vasishth (2022) argue that while regressions and rereading in eye tracking may be purely automatic to some extent, regressions in BSPR are always conscious, because the participant needs to make a deliberate decision to press the “back” key.

To test their hypothesis that selective reanalysis is connected to conscious awareness, Paape and Vasishth (2022) applied the BSPR paradigm to German garden-path sentences. At the disambiguating region, they found reading slowdowns consistent with garden-path effects in right-bounded reading times (that is, before readers read past the region), but no indication of selective rereading of earlier material. Instead, participants engaged in a variety of rereading strategies, including rereading of the entire sentence (see also von der Malsburg and Vasishth, 2011, 2013; Christianson et al., in prep.). Given the absence of evidence for selective reanalysis in a conscious reading setting, there is reason to doubt the assumption that selective rereading is tied to conscious awareness. However, a possible reason for the absence of selective rereading in Paape and Vasishth’s study is that the garden-path effects in the German sentences they used may be milder in comparison to the NP/Z ambiguity studied by Frazier and Rayner (1982).³ If

¹ Whether the comprehender actually engages in such deliberate reprocessing may also depend on their motivational state, as noted by Christianson et al. (2022). If the reader is not fully attentive or has otherwise mentally disengaged from the text, it is unlikely that they will make a deliberate effort to reanalyze the structure.

² See Kemtes et al. (2001) and Gong (2019) for similar approaches. Also see Hatfield (2016) for a related design using a touchscreen, and Vidal-Abarca et al. (2011) for a mouse-based design with larger text chunks.

³ Based on a corpus study, Sturt et al. (1999) report that 95% of comma-less NP/Z contexts are disambiguated towards the NP complement analysis. The zero complement analysis is thus 19 times less likely than the NP complement analysis in the English materials. By contrast, the estimated frequency ratio in favor of the preferred analysis in the materials of Paape and Vasishth (2022) was only 4.5:1.

the garden-path effect in the German materials was not strong enough to rise to consciousness, one would not expect conscious rereading to be triggered.

To further explore this issue, we conduct a BSPR study to investigate deliberate, selective rereading in NP/Z sentences, in which the garden path is often consciously registered. We also directly compare NP/Z sentences to the discourse inconsistency sentences used by Hessel et al. (2021), as well as to two additional constructions from the psycholinguistic literature. In order to distinguish confirmatory rereading in the BSPR task from revisionary rereading, we use end-of-trial acceptability judgments. This offline measure allows us to distinguish selective rereading that is purely confirmatory from selective reanalysis, which is revisionary by the definition of Frazier and Rayner (1982). Specifically, we ask participants to reject sentences if they believe they are ungrammatical or nonsensical. Garden-path sentences are sometimes rejected as ungrammatical, presumably because readers cannot find the correct syntactic analysis (Ferreira and Henderson, 1991; Warner and Glass, 1987). If selective rereading indexes selective reanalysis, and if reanalysis entails the computation of a grammatically well-formed structure for the sentence, participants should be more likely to judge garden-path sentences as grammatical in trials with selective rereading. By contrast, if readers merely reread to confirm their initial analysis, offline judgments should not vary depending on whether rereading occurred or not: Either participants succeed at reanalyzing the garden-path structure covertly, without regressing (Lewis, 1998), and then reread to confirm the reanalyzed structure, or rereading simply confirms the perceived ungrammaticality of the sentence.

Besides NP/Z garden-path sentences and discourse inconsistency sentences, our study uses sentences with a main clause/reduced relative clause (RRC) ambiguity (*the lawyer sent by the governor ...*), which causes milder garden-path effects than the NP/Z ambiguity. The fourth construction used in the study consists of reflexive sentences with grammatically unavailable but gender-matching distractor antecedents (*She remembered that the surgeon had pricked herself*; Sturt, 2003). These structures are theoretically interesting because they have been found to affect regression and rereading behavior in eye tracking, and are thus promising candidates for an investigation of conscious rereading. We present all four sentence types to the same participants in one experiment.

If the selective reanalysis hypothesis holds, and if selective reanalysis is a deliberate, conscious reading strategy, the BSPR paradigm should reveal revisionary rereading of critical sentence regions in NP/Z and RRC garden-path sentences that results in positive acceptability judgments. Discourse inconsistencies (*barks ... kitten*), on the other hand, should show confirmatory rereading with no effect on judgments, as there is no alternative structure or interpretation that can be derived. If participants miss the inconsistency on the first pass, judgments of discourse inconsistencies may even be negatively affected by rereading, as participants may become more certain that the discourse is indeed malformed. Finally, reflexive sentences may show both confirmatory and revisionary rereading, as we explain below.

To analyze first-pass reading and rereading in BSPR, one can compute reading measures that are familiar from the eye-tracking literature, such as first-pass reading times and rereading times. In addition to reporting such region-based measures, we also conduct a scanpath analysis. Scanpaths capture the trajectory of reading – usually in the form of eye fixations, but in our case in the form of key presses and reading times – across an entire trial. Different methods for comparing scanpaths between participants and conditions have been developed (see Anderson et al., 2015 for a review). We use the Scasim algorithm developed by Titus von der Malsburg,⁴ as it has previously been applied to garden-path sentences in eye-tracking and BSPR (Christianson et al., in prep.; von der Malsburg and Vasishth, 2011, 2013; Paape and Vasishth, 2022).

To ensure that our results replicate, we run two identical web-based BSPR studies with 100 participants each. We also run a standard SPR experiment in addition to the two BSPR experiments, in order to investigate the effects of the manipulations when rereading is not possible.

We now present the experimental setup and discuss each of the four manipulations in turn.

2. Materials

In order to increase the naturalness of reading and reduce participant fatigue, we use region-by-region as opposed to word-by-word SPR/BSPR. In all examples, the · symbol marks the boundaries between presentation chunks, that is, a key press is required at each · to reveal the next word or group of words.

2.1 NP/Z ambiguity

(3) The news show · stated · that · while · the team · $\left. \begin{array}{l} \text{trained,} \\ \text{trained} \end{array} \right\}$ · the striker · **wondered** ·
whether · the damage · would take long to heal.

The structure of the sentence is analogous to that of (2) above. In the condition with the comma, the end of the *while*-clause is overtly marked. By contrast, in the condition without the comma, *the striker* can initially be attached as the object of *trained*. Assuming that the parser prematurely makes this attachment, a garden-path effect is expected at *wondered* in the no-comma condition compared to the comma condition.

Our experiment uses 24 NP/Z sentences taken from the study of Mitchell et al. (2008). We use the versions with a late (as opposed to sentence-initial) NP/Z ambiguity, and with a short ambiguous region.

⁴ The R package is available at <https://github.com/tmalsburg/scanpath>; the algorithm is described in von der Malsburg and Vasishth (2011).

2.2 RRC ambiguity

- (4) Things · soon · turned ugly, · because · $\left\{ \begin{array}{l} \text{the lawyer} \\ \text{the package} \end{array} \right\}$ · sent · **by the governor** · was neglected · by the officials · in the town hall.

In the *lawyer* (animate) condition, the manipulated noun phrase can initially be plausibly analyzed as the subject of the verb *sent*. However, at the disambiguating phrase *by the governor*, it becomes clear that *sent* is actually passive and heads a reduced relative clause modifying *the lawyer*, which should result in a garden-path effect. In the *package* (inanimate) condition, the initial misanalysis is semantically implausible, which may reduce the garden-path effect compared to the animate condition.

The 24 RRC sentences used in our study are based on the materials of Clifton et al. (2003). We added a longer preamble, as in (4), so that the critical RRC structure does not appear at the beginning of the sentence.⁵

2.3 Discourse inconsistency

- (5) Rover · barks · at all passing animals · on the street. · He's · $\left\{ \begin{array}{l} \text{the most alert puppy} \\ \text{the most alert kitten} \end{array} \right\}$ · in the whole neighborhood, · in fact.

The manipulation of discourse (in)consistency is very straightforward: In the *puppy* (consistent) condition, the unfolding discourse is plausible. By contrast, in the *kitten* (inconsistent) condition, the discourse is implausible, as kittens do not bark.

Our study uses 20 discourse-inconsistency items taken from Connor et al. (2015).

2.4 Similarity-based interference in reflexives

- (6) $\left\{ \begin{array}{l} \text{Jonathan} \\ \text{Jennifer} \end{array} \right\}$ · was pretty worried · at the City Hospital. · $\left\{ \begin{array}{l} \text{He} \\ \text{She} \end{array} \right\}$ · remembered · that · the surgeon · **had pricked herself** · badly · with a used syringe needle.

In the *Jonathan* (distractor mismatch) condition, the gender of the reflexive *herself* mismatches the gender of the initial noun phrase (the distractor) and its coreferential pronoun *He*. Despite

⁵ Unlike in the study of Clifton et al., all sentences also had the same ending across conditions, and each sentence was seen only once by each participant.

the NP *the surgeon* being biased towards a male referent, it should thus be clear that the reflexive *herself* can only refer to the surgeon. By contrast, in the *Jennifer/She* (distractor match) condition, the reflexive may be taken to refer to *Jennifer/She* instead, which would violate grammar, specifically Principle A of binding theory (Chomsky, 1981).

Our study uses 24 reflexive items taken from the accessible-mismatch conditions of Sturt's (2003) Experiment 1. As in the original materials, half of the sentences use *himself* as the pronoun while the other half use *herself*. The gender of the reflexive mismatches the prototypical gender of the antecedent NP in all sentences. The materials were adapted from British to American English, including both spelling and vocabulary differences ("*footballer*"). Some of the first names (*Jennifer/Jonathan*) appearing in the sentences were also adapted to reflect a greater variety of cultural backgrounds, that is, names like *Rajesh*, *Fatima* and *Daewon* were included.

3. Previous findings and predictions

For ease of reference, **Table 1** gives an overview of all experimental manipulations.

Table 1: Overview of the experimental manipulations. Manipulated regions are underlined, critical regions appear in bold.

NP/Z garden path	
... while the team <u>trained</u> , the striker wondered ...	Comma
... while the team <u>trained</u> the striker wondered ...	No comma
RRC garden path	
... because <u>the lawyer</u> sent by the governor was neglected ...	Animate NP
... because <u>the package</u> sent by the governor was neglected ...	Inanimate NP
Discourse inconsistency	
Rover barks at all passing animals ... He's <u>the most alert puppy</u> ...	Consistent
Rover barks at all passing animals ... He's <u>the most alert kitten</u> ...	Inconsistent
Reflexive	
<u>He</u> remembered that the surgeon had pricked herself ...	Mismatching distractor
<u>She</u> remembered that the surgeon had pricked herself ...	Matching distractor

3.1 NP/Z sentences

NP/Z sentences are often judged to be ungrammatical when the comma is absent (e.g., Warner and Glass, 1987; Ferreira and Henderson, 1991, 1993), and show evidence of increased processing difficulty in the eye-tracking record in the form of longer (re)reading times and regressions (e.g., Frazier and Rayner, 1982; Mitchell et al. 2008; Schotter et al. 2014).

The selective reanalysis hypothesis predicts more regressions from the disambiguating verb when the comma is absent compared to when it is present, as well as more extensive rereading of the embedded ambiguous verb (*trained*), the main clause subject (*the striker*), and possibly the disambiguating main clause verb (*wondered*). Under the assumption that selective rereading is revisionary, more positive acceptability judgments should be given in trials in which the critical sentence regions are reread, compared to trials in which they are not. By contrast, if rereading is merely confirmatory, judgments should not be affected by the presence or absence of rereading, or should even be negatively affected by rereading, assuming that subjects become more certain that no correct analysis is possible.

3.2 RRC sentences

There has been some controversy in the literature with regard to whether the animacy manipulation in RRC sentences (*the lawyer/package sent ...*) can affect first-pass parsing in the sense that the RRC analysis will be preferred over the main clause analysis in the inanimate condition, or whether it only becomes relevant during reanalysis (Ferreira and Clifton, 1986; Rayner et al., 1983; Trueswell et al., 1994; Clifton et al., 2003; Zhang and Witzel, 2021). It is not our aim to resolve this debate here. Rather, we ask whether, *if* there is reanalysis, and if it is affected by the animacy manipulation, such an effect will be visible in measures of selective rereading.

If selective reanalysis does occur in RRC sentences, one would expect additional regressions from the *by*-phrase in sentences with animate NPs compared to sentences with inanimate NPs, as well as more extensive rereading of the manipulated NP (*the lawyer/the package*) and the following ambiguous verb (*sent*). As for NP/Z sentences, selective rereading should positively affect judgments if it is revisionary, and should not, or negatively, affect judgments if it is confirmatory.

3.3 Discourse inconsistency sentences

In eye-tracking studies with children, inconsistent discourse continuations compared to consistent discourse continuations have been found to lead to longer first-pass reading times, as well as more extensive rereading of the critical *puppy/kitten* region (Connor et al., 2015; Zargar et al., 2020; Hessel et al., 2021), particularly in children with stronger reading skills.

If rereading in discourse inconsistency sentences is mostly automatic, as suggested by Hessel et al. (2021), adults should, in principle, show the same reading patterns as children, that is, selective rereading of inconsistent continuations. However, the more conscious nature of the BSPR paradigm could alter the rereading patterns. In light of the BSPR results of Paape and Vasishth (2022), participants may show more unselective, whole-sentence rereading as they try to make sense of the incongruous discourse. As revisionary rereading is not expected to occur in sentences with inconsistent discourse continuations, there should be no positive effect

of rereading on their acceptability. Confirmatory rereading, on the other hand, could lead to decreased acceptability if subjects sometimes miss the inconsistency on the first pass.

3.4 Reflexive sentences

In Sturt's (2003) eye-tracking study, the distractor match condition showed more first-pass regressions, longer regression paths and longer second-pass reading times on the reflexive *herself*. Other studies have found the opposite pattern, with distractor-match conditions being easier to process, while still others found no significant differences between conditions (see Jäger et al., 2017 for a review).

In BSPR, selective rereading in reflexive sentences may be confirmatory in the sense that participants will want to make sure that a) the referent of the NP *the surgeon* is really female, contrary to their expectation, and/or that b) the structure of the sentence is really such that the reflexive cannot refer to the distractor. A more radical proposal would be that rereading can also be revisionary: Participants could reanalyze the reflexive as referring to the distractor on the second pass in the distractor-match condition. Under this account, one would expect more positive acceptability judgments in trials with rereading in the distractor-match condition, compared to trials without rereading.

4. Participants

For each of the two identical BSPR experiments (BSPR1 and BSPR2) and the standard SPR experiment, 100 self-identified native speakers of English were recruited over Prolific (<https://www.prolific.co>; Palan and Schitter, 2018). All participants indicated that they were currently living in the US. They received £ 3.50 for their participation.⁶

5. Procedure

The three experiments were run on the Ixcel farm (<https://spellout.net/ixcelfarm>; Drummond, 2013). Within each experiment, every participant read 92 sentences in total (24 NP/Z + 24 RRC + 20 Discourse + 24 Reflexive). Sentences were rotated through the experimental conditions according to a Latin-square design, and presentation order was randomized. Due to the amount and varied nature of the experimental materials, no fillers were used. Participants gave informed consent after reading the instructions, then completed two practice trials, then proceeded to the main experiment.

The procedure was the same as in Paape and Vasishth (2022), apart from the nature of the end-of-trial task: Paape and Vasishth asked comprehension questions and rewarded accuracy and

⁶ Even though we targeted American English speakers, the Prolific platform is based in the UK, and payment is thus calculated in British Pounds.

reading speed with bonus payouts, whereas we collected acceptability judgments and offered no additional rewards. We chose acceptability judgments as our dependent measure because we were interested in the connection between selective rereading in a given trial and the end-of-trial judgment.

The instructions for the judgments were as follows:

When judging the acceptability of the sentences, please take into account both meaning and grammar. Sentences should be rejected as unacceptable if they violate the rules of English grammar (“Peter goed swimming yesterday”) and/or if their meaning is contradictory or implausible (“Mary ate a table for dinner yesterday”). Possible stylistic issues should be disregarded.

After each sentence, participants were presented with the question “Was the sentence acceptable?”. The possible responses were “Yes”, “No”, and “I don’t know”.

Because we were interested in the subjective naturalness of the BSPR paradigm compared to standard SPR, participants were presented with a debriefing screen after all sentences had been read and judged. Both in the BSPR studies and in the SPR study, participants were asked “Did reading feel natural in this experiment?” and “Were the sentences easy to understand?”. Only in the BSPR experiments, they were also asked “Was rereading helpful in this experiment?”. For each question, participants had four answer options: “Yes, very”, “Mostly yes”, “Mostly not”, and “Not at all”. Participants were also given the opportunity to comment on the experiment using a text box.

6. Data analysis

Data from seven participants (three in BSPR1, four in BSPR2) were removed because the participants either skipped a majority of sentences by using the CTRL key or indicated during debriefing that they had done an experiment with similar sentences before. From the remaining data, we removed all trials in which any single-region reading time was longer than 5000 ms or in which any region had not been read at all. This resulted in a loss of 7% of the data in both BSPR experiments and 6% of the data in the SPR experiment.

For every trial, we coded whether the acceptability judgment was positive or negative and, for BSPR only, whether the trial contained at least one regression. For each experiment, we fitted linear mixed-effects models (LMMs) with random slopes by participants and by items using the *brms* package for Bayesian inference (Bürkner, 2017, 2018) in R (R Core Team, 2020). Four chains with 4000 iterations each were run for all models. Acceptability judgments were fitted with a Bernoulli likelihood, using the factor condition as a predictor. The factor was sum-coded as follows:

- NP/Z ambiguity: Comma -1, No-comma +1
- RRC ambiguity: Animate NP -1, Inanimate NP +1
- Discourse inconsistency: Consistent continuation -1, Inconsistent continuation +1
- Reflexives: Mismatching distractor -1, Matching distractor +1

For the BSPR experiments, the overall proportion of regressions at the trial level (0 = no regression, 1 = at least one regression) was analyzed using a Bernoulli likelihood. In addition to the fixed effect of condition, the model contained the acceptability judgment at the end of the trial as a sum-coded predictor (unacceptable -1, acceptable +1), as well as the interaction between condition and judgment.

Our region-wise analysis of BSPR reading measures is focused on first-pass parsing and selective rereading, and closely follows that of Paape and Vasishth (2022). We selected a relatively small number of reading measures, given that each additional measure analyzed increases the risk of false-positive findings (von der Malsburg and Angele, 2017). We chose right-bounded reading time (RBRT), regressive rereading time (RegRRT), and the proportion of first-pass regressions from a region as our measures of interest. Right-bounded reading time is the sum of all reading times on the region before moving on to the next region; it is composed of first-pass reading time and progressive rereading time. We define regressive rereading time as the sum of all non-zero reading times on the region after having moved on to the next region, including both regressive and progressive revisits: After having read region $n + 1$, participants may either return to region n from the right, or jump back to the beginning of the sentence and return to region n from the left; both types of revisits count towards regressive rereading time for region n . Due to the exclusion of regions with $\text{RegRRT} = 0$ from the analysis, the measure is deconfounded from rereading probability (Leininger et al., 2017; Vasishth and Drenhaus, 2011), and the data can be analyzed without violating distributional assumptions. Informative differences in rereading probability between conditions are captured by the scanpath analysis reported below.

The RBRT and RegRRT measures were analyzed using LMMs with a Lognormal likelihood. Following Paape and Vasishth (2022), we assume that data points below 150 ms in BSPR are equivalent to skipping a region. Removing these data points resulted in a loss of 2%/<1% of the data for RBRT and 84%/85% of the data for RegRRT, counting removed data points with $\text{RegRRT} = 0$. For both reading measures, data points above 5000 ms were treated as outliers and were also removed from the analyses, resulting in the loss of less than 1% of the data in both experiments.

For NP/Z, RRC and discourse inconsistency sentences, we defined a window of analysis that contained the critical region as well as the two regions to the left and one region to the right. The critical region in NP/Z sentences is the disambiguating verb, that is, *wondered* in (3). For RRC sentences, the critical region is the ambiguous participle, that is, *sent* in (4). For discourse

inconsistency sentences, the critical region is the region containing the manipulated NP, that is *the most alert puppy/kitten* in (5).⁷ Finally, for reflexive sentences, the critical region is the region containing the reflexive, that is, *had pricked herself* in (6).

For all sentence types except RRC sentences, the sentence ended with the second region after the critical region in most (but not all) items. When the second-next region from the critical region was *not* the sentence-final region, the final region was analyzed instead. For reflexive sentences, we did not analyze reading times at the complementizer *that*, but instead included the pronoun *She/He* in the window of analysis, as readers could plausibly reread this region after encountering the matching or mismatching reflexive.

Rather than using a binary criterion to distinguish “present” from “absent” effects in the reading measures, we focus on quantifying uncertainty by presenting the estimates from our analyses with their accompanying uncertainty intervals (Vasishth and Gelman, 2021). When interpreting the results, the reader should consider the width of the credible intervals, and whether the effect is consistent across the experiments. The full range of results is given in the online supplement at <https://osf.io/54bda>.

7. Results – Debriefing

We recoded the debriefing responses numerically (“Yes, very” = 4, “Mostly yes” = 3, “Mostly not” = 2, “Not at all” = 1) and computed the mean ratings for each question for each method (BSPR vs SPR). BSPR was judged to be somewhat more natural than SPR (2.58 vs 2.34), and sentences were also judged to be somewhat more comprehensible in BSPR than in SPR (2.45 vs 2.24). A Bayes factor analysis shows anecdotal evidence of BSPR being considered more natural than standard SPR (BF = 2.23), as well as moderate evidence of BSPR being easier to comprehend than standard SPR (BF = 8.66). Rereading in BSPR was judged to be highly helpful (mean 3.38).

8. Results – Acceptability judgments and regressions

Table 2 shows the proportion of positive acceptability judgments, as well as trial-level regression probabilities (BSPR only) by experiment, sentence type and condition.

Figure 1a shows LMM estimates for the acceptability judgments. Acceptability judgments were reliably affected by the manipulations across the three experiments: NP/Z sentences without commas were accepted far less often than NP/Z sentences with commas. Sentences with inconsistent discourse continuations were accepted far less often than sentences with consistent

⁷ Discourse inconsistency sentences are the only sentence type for which the critical region contained different words in the different conditions. While the lexical difference may affect reading times at the critical region, Connor et al. (2015) matched the words across the two conditions for word length, morphological complexity, and frequency, thus alleviating the concern.

Table 2: Proportion of positive acceptability judgments $p(\text{pos})$ and proportion of regressions $p(\text{reg})$ by experiment, sentence type and condition.

Type	Condition	$p(\text{pos})$	$p(\text{pos})$	$p(\text{pos})$	$p(\text{reg})$	$p(\text{reg})$
		BSPR1	BSPR2	SPR	BSPR1	BSPR2
NP/Z ambiguity	Comma	0.80	0.84	0.80	0.32	0.29
NP/Z ambiguity	No comma	0.40	0.39	0.43	0.41	0.39
RRC ambiguity	Animate	0.79	0.80	0.74	0.32	0.34
RRC ambiguity	Inanimate	0.87	0.88	0.87	0.27	0.25
Discourse	Consistent	0.88	0.92	0.89	0.25	0.24
Discourse	Inconsistent	0.49	0.52	0.52	0.31	0.29
Reflexives	Mismatch	0.78	0.79	0.75	0.33	0.32
Reflexives	Match	0.70	0.72	0.72	0.33	0.32

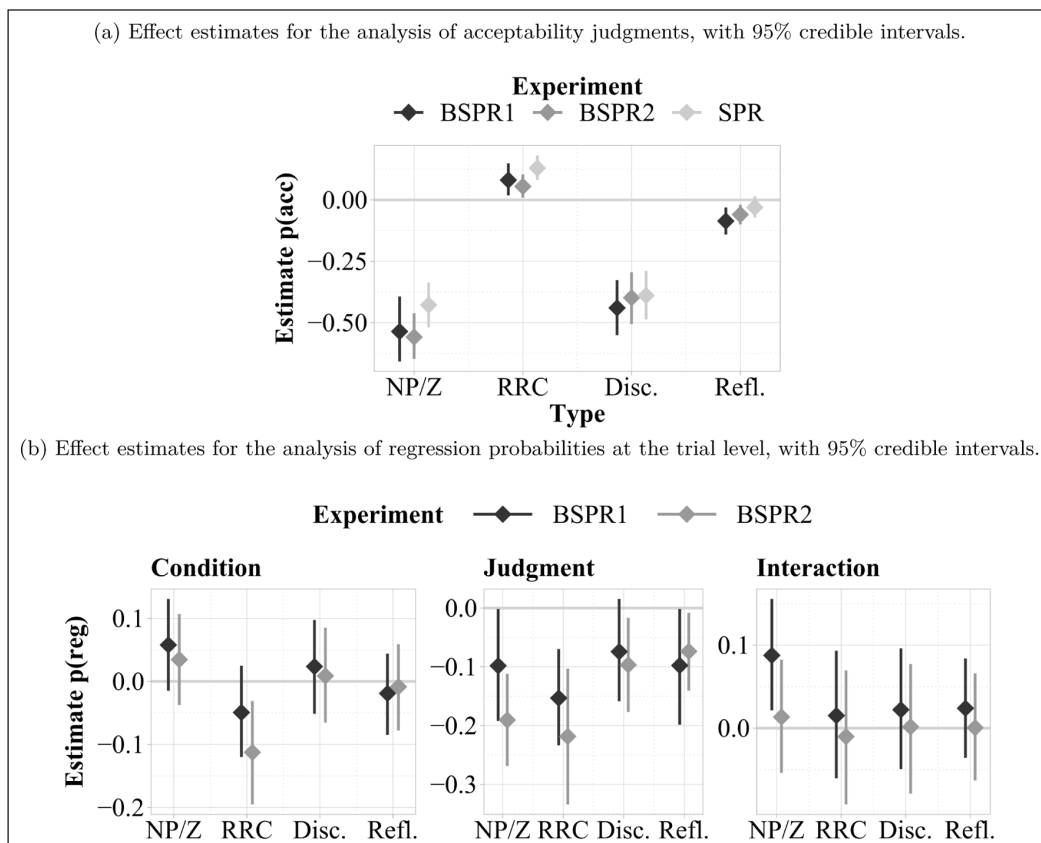


Figure 1: Effect estimates from the LMM analyses of acceptability judgments and regression probabilities.

discourse continuations. RRC sentences with inanimate NPs were accepted more often than RRC sentences with animate NPs, though the difference was much smaller in magnitude compared to NP/Z and discourse inconsistency sentences. Finally, reflexive sentences with matching distractors were accepted less often than reflexive sentences with mismatching distractors, but the difference between conditions was relatively small.

Figure 1b shows LMM estimates for regression probabilities at the trial level. Across all sentence types, BSPR regressions occurred less often in trials with positive acceptability judgments compared to trials with negative acceptability judgments. NP/Z sentences without commas showed more regressions than NP/Z sentences with commas, while RRC sentences with inanimate NPs showed fewer regressions than RRC sentences with animate NPs. In BSPR1, NP/Z sentences showed an interaction, such that trials with positive judgments only showed fewer regressions in the comma condition, but there was no indication of such an effect in BSPR2. Across both experiments, discourse inconsistency sentences and reflexive sentences showed no indication that the manipulations affected regression probabilities at the trial level, despite a numerical tendency in discourse inconsistency sentences towards inconsistent continuations leading to more regressions.

9. Results – Reading measures

9.1 NP/Z sentences

Reading time measures for NP/Z sentences by experiment and sentence region are shown in **Figure 2a** (BSPR only). **Figure 2b** shows the estimates of the region-wise LMM analyses. Estimates from the standard SPR experiment are shown next to the RBRT results from the BSPR experiments. For a plot showing reading time measures by region for the SPR experiment, please refer to the online supplementary materials.

Despite some differences in the distribution of effects, the results converge across experiments: Readers slowed down when they encountered a comma after the ambiguous verb (*trained(,)*), that is, when the clause boundary was explicitly marked. This slowdown is presumably the result of being forced to adopt the dispreferred intransitive reading of the verb, or may simply be a low-level effect of processing punctuation (Warren et al., 2009). At the disambiguating main clause verb (*wondered*), a garden-path effect was visible in right-bounded BSPR reading times in the no-comma compared to the comma conditions, which in BSPR2 continued into the following spillover region. The garden-path effect was also visible in first-pass regressions at the point of disambiguation (*wondered*) and the two following regions. Regressive rereading times were elevated in the no-comma conditions in the NP region (*the striker*) and the disambiguating verb region. As **Figure 2b** shows, right-bounded reading times in BSPR closely align with reading times in standard SPR.

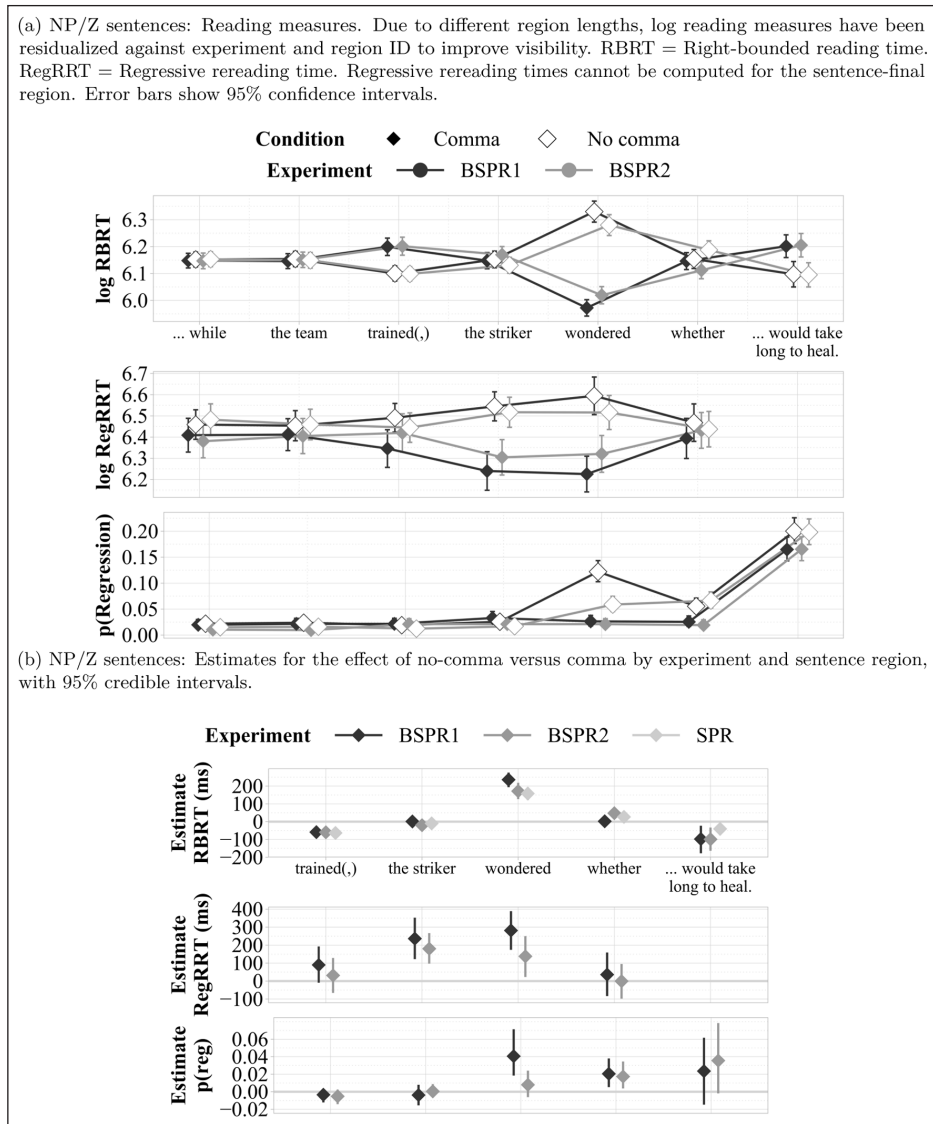


Figure 2: Reading measures and estimates from the LMM analyses for NP/Z sentences.

9.2 RRC sentences

Reading time measures for RRC sentences by experiment and sentence region are shown in Figure 3a (BSPR only). Figure 3b shows the estimates of the region-wise LMM analyses. Results from the standard SPR experiment are shown next to the RBRT results from BSPR.

The results for RRC sentences also largely converge across the two BSPR experiments: At the *by*-phrase and in the following region, participants read more slowly in sentences with animate NPs compared to sentences with inanimate NPs, consistent with a stronger garden-path effect. Right-bounded reading times in BSPR were again closely aligned with reading times in standard SPR.

In BSPR2 and SPR, there was some indication that readers slowed down at the verb (*sent*) when the preceding NP was inanimate, which contradicts the preferred active interpretation. Mainly

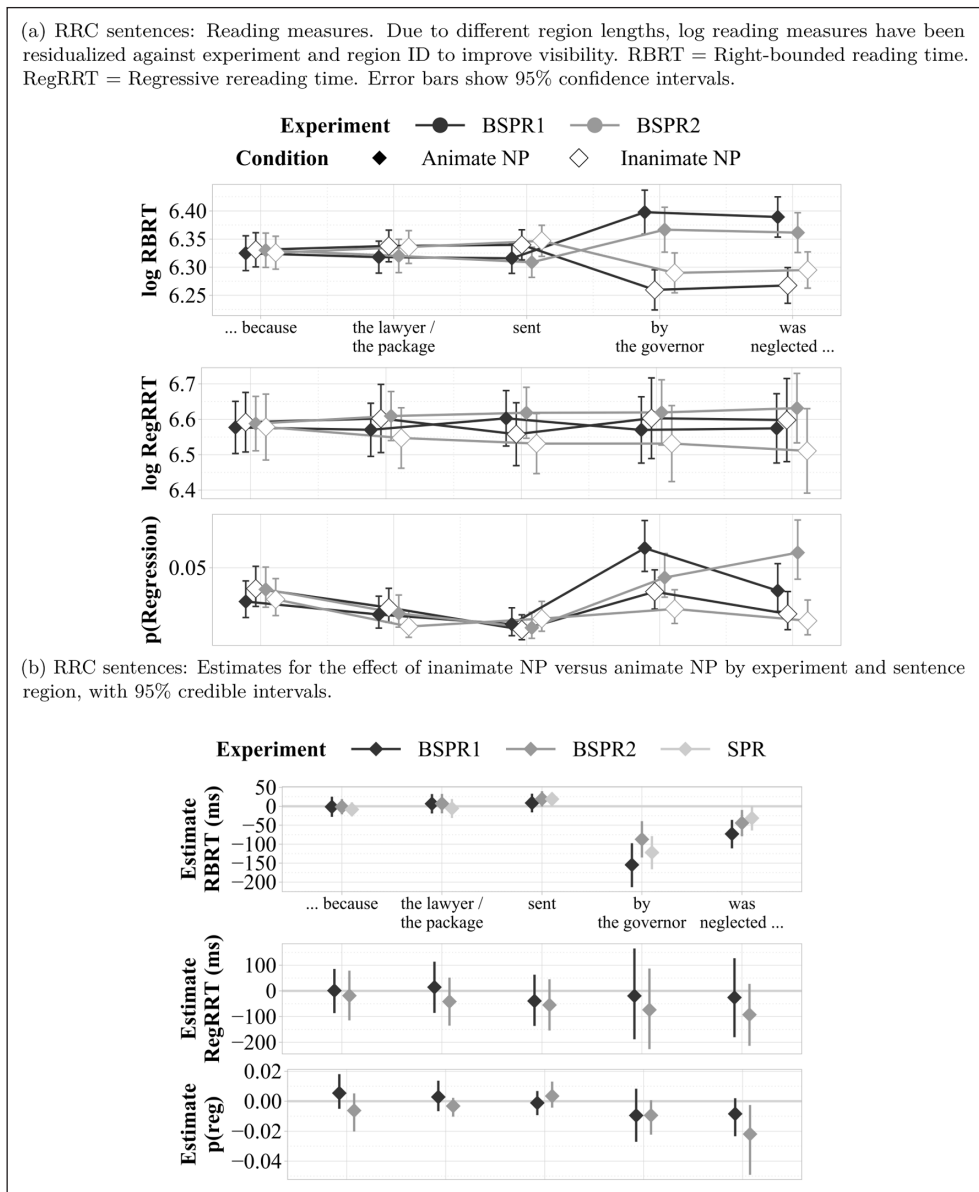


Figure 3: Reading measures and estimates from the LMM analyses for RRC sentences.

in BSPR2, there was also some indication that stronger garden-pathing at the *by*-phrase in the animate-NP condition led to more regressions. A numerical difference between the experiments was visible across the entire window of analysis in regressive rereading times: In BSPR2, all theoretically relevant regions tended to be reread more in the animate-NP condition, while no such pattern was visible in BSPR1. Speculatively, this mismatch between the experiments may be due to different preferred reading strategies in the two samples of participants: As noted by Paape and Vasishth (2022) and Christianson et al. (in prep.), rereading strategies in response to garden-pathing may differ widely between individuals. Our scanpath analyses reported below shed additional light on the range of these differences. Alternatively, the differences between

experiments may simply be due to noise in the rereading data, given that there were much fewer data points compared to the first-pass reading data.

9.3 Discourse inconsistency sentences

Reading time measures for discourse inconsistency sentences by experiment and sentence region are shown in **Figure 4a** (BSPR only). **Figure 4b** shows the estimates of the region-wise LMM analyses. Results from the standard SPR experiment are shown next to the RBRT results from BSPR.

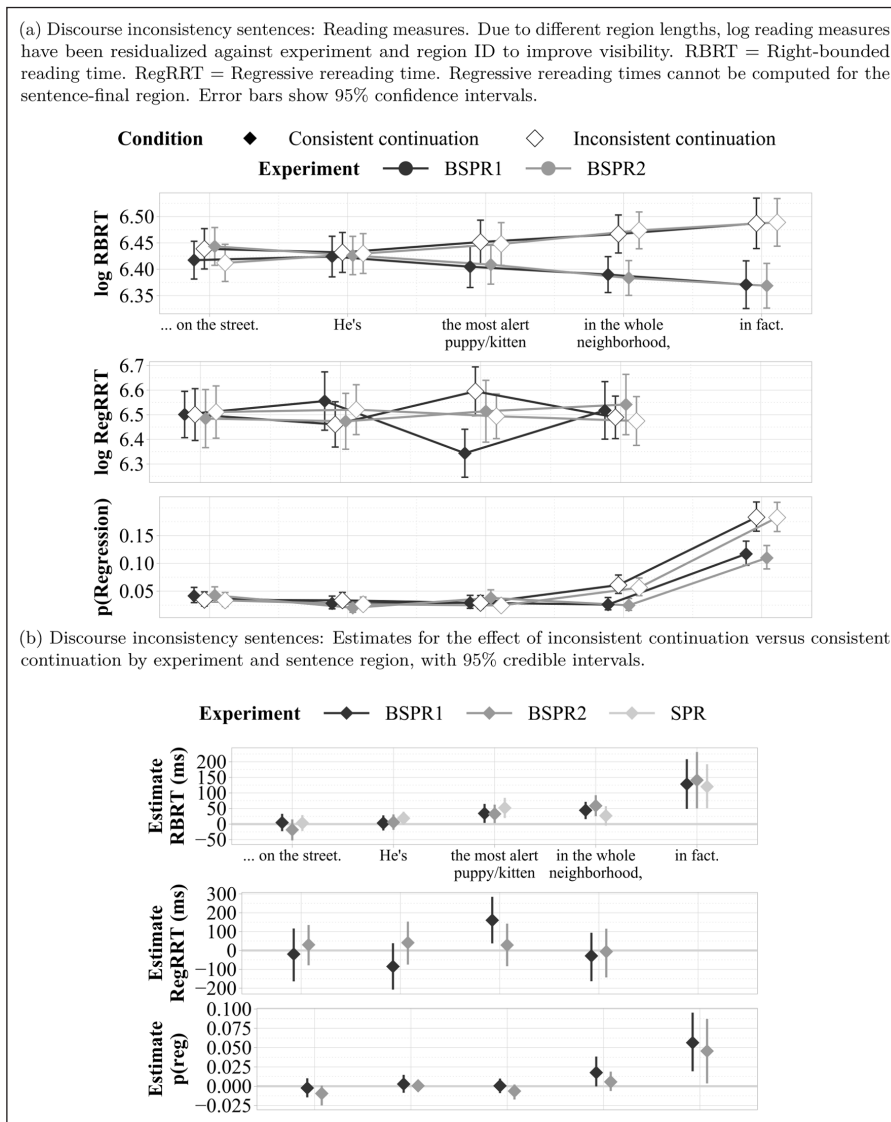


Figure 4: Reading measures and estimates from the LMM analyses for discourse inconsistency sentences.

As for the garden-path sentences, the results for discourse inconsistency sentences mostly align across experiments: Encountering an inconsistent discourse continuation as opposed to a consistent discourse continuation led to longer right-bounded reading times in the manipulated region and the following regions, as well as more regressions. Right-bounded reading times in BSPR aligned with reading times in standard SPR.

One pattern that was very different across the two BSPR experiments was the effect of the manipulation on regressive rereading times at the critical region: BSPR1 showed a clear effect that is in line with selective rereading of inconsistent continuations, matching the findings of Connor et al. (2015), Zargar et al. (2020), and Hessel et al. (2021) in children. By contrast, BSPR2 showed no indication of such an effect. As for RRC sentences, this mismatch between the experiments may speculatively be due to different preferred reading strategies in the two participant samples or due to noise in the data.

9.4 Reflexive sentences

Reading time measures for reflexive sentences by experiment and sentence region are shown in **Figure 5a** (BSPR only). **Figure 5b** shows the estimates of the region-wise LMM analyses. Results from the standard SPR experiment are shown next to the RBRT results from BSPR.

The results for reflexive sentences can be summarized quickly: There was no indication that any of the online reading measures in the theoretically interesting regions were affected by the manipulation of distractor match/mismatch. The numerical patterns observed in BSPR1 that could be of potential theoretical interest, that is, the patterns in right-bounded and regressive reading times of the critical reflexive region, were reversed in BSPR2. The only region that showed a consistent effect across the BSPR experiments was the final region of the sentence: The distractor-match conditions were more difficult to process at this point than the distractor-mismatch conditions. This effect is likely due to end-of-sentence wrap-up, which “attempt[s] to handle any inconsistencies that could not be resolved during the sentence” (Just and Carpenter, 1980, p. 345). Given that there was no indication of such an effect in standard SPR, it may be driven by progressive rereading.

10. Discussion

The results for NP/Z sentences indicate a very robust garden-path effect. Crucially, the garden-path effect was visible in first-pass regressions from the disambiguating region, as well as in regressive rereading times for the disambiguating region and the ambiguous NP. This is in line with the selective reanalysis hypothesis of Frazier and Rayner (1982), as well as with the proposal of Paape and Vasishth (2022) that selective reanalysis is conscious.

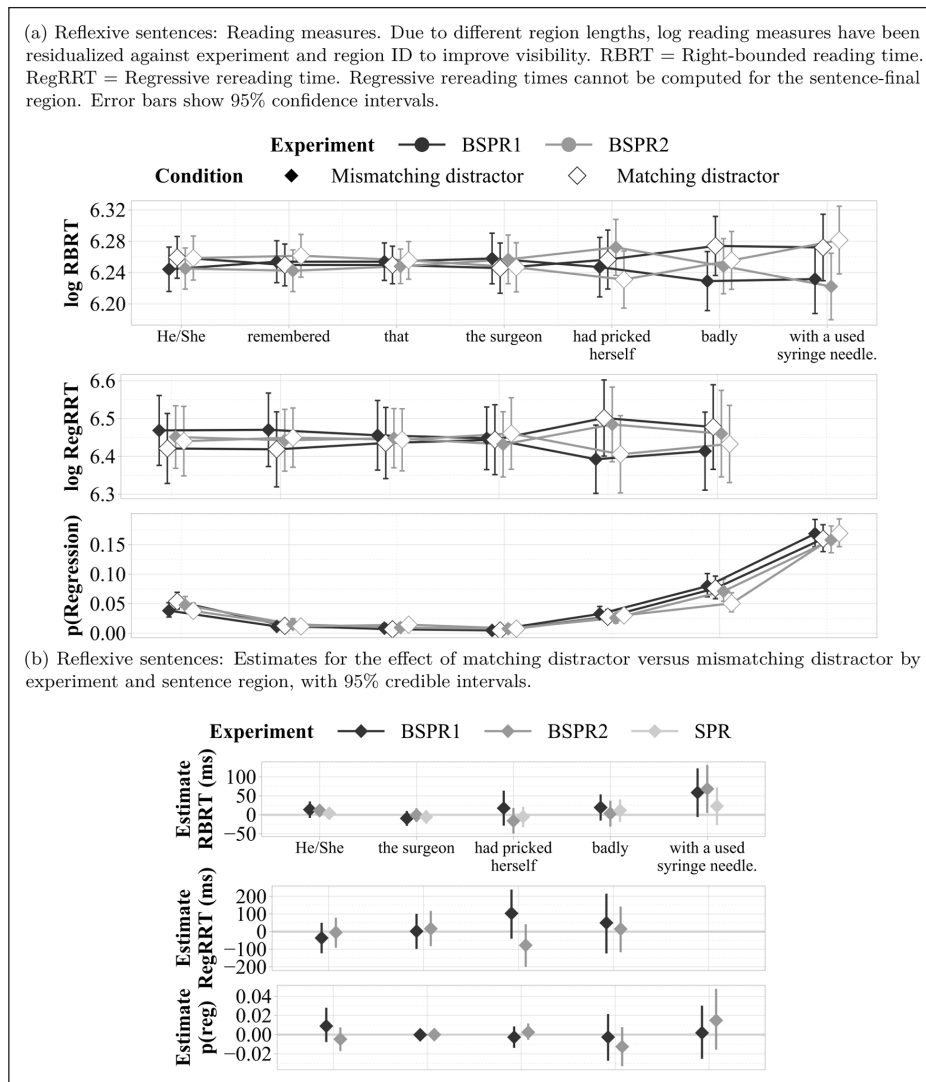


Figure 5: Reading measures and estimates from the LMM analyses for reflexive sentences.

However, the large proportion of rejections in the offline judgments casts doubt on the selective reanalysis hypothesis: Assuming that selective reanalysis is usually successful and leads to positive acceptability judgments, one would have expected more than 40% positive judgments for NP/Z sentences, and rereading should have correlated with higher acceptability. The relatively low proportion of positive judgments suggests that readers engaged in selective rereading for confirmatory rather than for revisionary purposes. That is, they reread earlier regions mainly to check their syntactic analysis and/or their interpretation but did not necessarily revise it. We will return to this point after presenting the scanpath results.

Compared to NP/Z sentences, the differences between conditions were smaller for RRC sentences, and regressive rereading was not consistently affected by the animacy manipulation

across the two BSPR experiments. Based on the end-of-trial judgment results, it appears that the NP/Z garden path is much stronger than the RRC garden path, which would be in line with previous findings (Van Schijndel and Linzen, 2021). NP/Z sentences without commas were rejected 60% of the time while RRC sentences with animate NPs were rejected only 20% of the time. It is thus possible that conscious, selective rereading is linked to the strength of the garden path as reflected in offline judgments: Garden paths that lead to more rejections may also show more selective rereading in general.

Discourse inconsistency sentences showed an interesting mismatch between online and offline measures compared to NP/Z sentences: The two manipulations had similar effects on the end-of-trial acceptability judgments, but the effect on reading times was much smaller for discourse inconsistency sentences than for NP/Z sentences, and there was no consistent evidence for selective rereading. Speculatively, this could be because participants sometimes outright fail to parse NP/Z sentences, while discourse inconsistencies can be parsed but are rejected as implausible. Compared to NP/Z sentences, readers need to consult their situation model in order to correctly accept or reject discourse inconsistency sentences (Hessel et al., 2021; Zargar et al., 2020; Connor et al., 2015; see van Moort et al., 2018 for a review), which may lead to more subtle effects compared to perceived ungrammaticality.

Finally, the results for reflexive sentences are compatible with the assumption that similarity-based interference affects end-of-sentence wrap-up, but does not cause conscious confirmatory or revisionary rereading during online processing. Given the mixed results in the literature for this construction (e.g., Dillon et al., 2013; Jäger et al., 2017; Parker and Phillips, 2017), it is not surprising that reflexives showed the weakest effects out of the four constructions tested.

11. Scanpath analysis of BSPR data

Scanpaths cover the trajectory of reading across the entire trial, and thus offer a more holistic view of the reading process than region-based measures. Scanpath analyses can uncover reading patterns that traditional region-based analyses may miss, including selective rereading (von der Malsburg and Vasishth, 2011; 2013; Christianson et al., in prep.; Paape and Vasishth, 2022).

We analyzed scanpaths for each sentence type separately by applying the Scasim algorithm for eye-tracking data (<https://github.com/tmalsburg/scanpath>; von der Malsburg and Vasishth, 2011) to the combined data from both BSPR experiments. Given that the total number of regions differed between individual sentences, we selected a subset of theoretically interesting regions for the scanpath analysis (see **Figures 6** and **7** below) and dropped all visits to other regions from the data. Furthermore, for each trial, we considered only the part of the overall scanpath following the first visit to the critical region. In order to capture whole-sentence rereading, which is often observed in garden-path sentences (von der Malsburg and Vasishth, 2011, 2013; Paape and Vasishth, 2022), we included the first and last regions of the sentence in the analysis across

all constructions. For the discourse inconsistency stimuli, we also added the inconsistent word in the first sentence (*barks ... kitten*) to the analysis, as readers may regress to this word in order to check their interpretation.

Following Paape and Vasishth (2022), we first simulated horizontal eye movements based on the BSPR data by using region IDs as x-coordinates while keeping the y-coordinate constant, and computed scanpath similarities with Scasim. We then applied nonmetric multidimensional scaling (Kruskal, 1964a, b; Oksanen et al., 2020) to map the similarities into a 2-dimensional space, and finally used model-based clustering (Scrucca et al., 2016) to identify clusters of similar scanpaths. Please refer to Paape and Vasishth (2022) for further details regarding the procedure.

11.1 Results

For reasons of space, we only report the descriptive results of the scanpath analysis. **Table 3** shows the proportion of trials belonging to each cluster in each condition, as well as the proportion of positive acceptability judgments within each cluster. **Figures 6** and **7** show example scanpaths from each identified cluster for each sentence type. We also conducted logistic regression analyses of cluster membership by condition as well as positive acceptability judgments by cluster. The results are given in the supplementary materials.

One shared type of scanpath cluster – which we will refer to as a meta-cluster – was consistently identified across all four sentence types, namely a cluster in which the critical portion of the sentence was only read once. We call this meta-cluster the “no rereading” cluster.⁸ Across all sentence types, trials were less likely to belong to this meta-cluster in the more difficult condition (no comma, animate NP, inconsistent discourse continuation, matching distractor). A second meta-cluster emerged for all sentence types apart from the RRC sentences: In this cluster, participants continued reading forward until the end of the sentence after reading the critical region, then returned to the beginning of the sentence in a region-by-region fashion and reread the entire sentence. We call this meta-cluster the “reread from beginning” cluster. Trials in this cluster were equally likely to appear in the more difficult and the less difficult conditions. A third meta-cluster that was common to all sentence types absorbed all reading patterns that did not fit into any of the other clusters, including single-region regressions at various points of the sentence, uses of the dedicated “back to beginning” (ESC) or “forward to judgment” (CTRL) keys, “zig-zag” reading patterns, and various other shapes. We call this cluster the “varied rereading” cluster.

⁸ For RRC sentences, the algorithm identified two “no rereading” clusters, but as there were no discernible differences between them, we merged them into one.

Table 3: Proportion of scanpaths belonging to each cluster p(tr), and proportion of positive acceptability judgments within each cluster p(pos) by sentence type and condition. Note that the “no rereading” proportions do not necessarily match the numbers given in Table 2, given that some regions were excluded from the scanpath analysis. The numbers in parentheses indicate the number of participants (out of 193 total) who used the respective strategy at least once.

NP/Z sentences										
	No rereading		Reread from beginning		Varied rereading		Selective rereading			
	(192)		(107)		(162)		(57)			
	p(tr)	p(pos)	p(tr)	p(pos)	p(tr)	p(pos)	p(tr)	p(pos)		
Comma	0.75	0.85	0.06	0.72	0.18	0.72	0.01	0.79		
No-comma	0.64	0.45	0.06	0.45	0.26	0.27	0.04	0.25		
RRC sentences										
	No rereading		Varied rereading		Local difficulty					
	(191)		(153)		(59)					
	p(tr)	p(pos)	p(tr)	p(pos)	p(tr)	p(pos)				
Animate NP	0.71	0.82	0.26	0.70	0.03	0.67				
Inanimate NP	0.81	0.91	0.17	0.81	0.02	0.94				
Discourse inconsistency sentences										
	No rereading		Reread from beginning		Varied rereading					
	(191)		(63)		(160)					
	p(tr)	p(pos)	p(tr)	p(pos)	p(tr)	p(pos)				
Consistent	0.83	0.91	0.03	0.80	0.15	0.84				
Inconsistent	0.75	0.54	0.04	0.40	0.21	0.41				
Reflexive sentences										
	No rereading		Reread from beginning		Varied rereading		Reread reflexive		Local rereading	
	(191)		(38)		(139)		(65)		(38)	
	p(tr)	p(pos)	p(tr)	p(pos)	p(tr)	p(pos)	p(tr)	p(pos)	p(tr)	p(pos)
Mismatch	0.77	0.81	0.03	0.87	0.14	0.72	0.03	0.82	0.02	0.64
Match	0.77	0.75	0.02	0.88	0.16	0.60	0.04	0.81	0.02	0.50

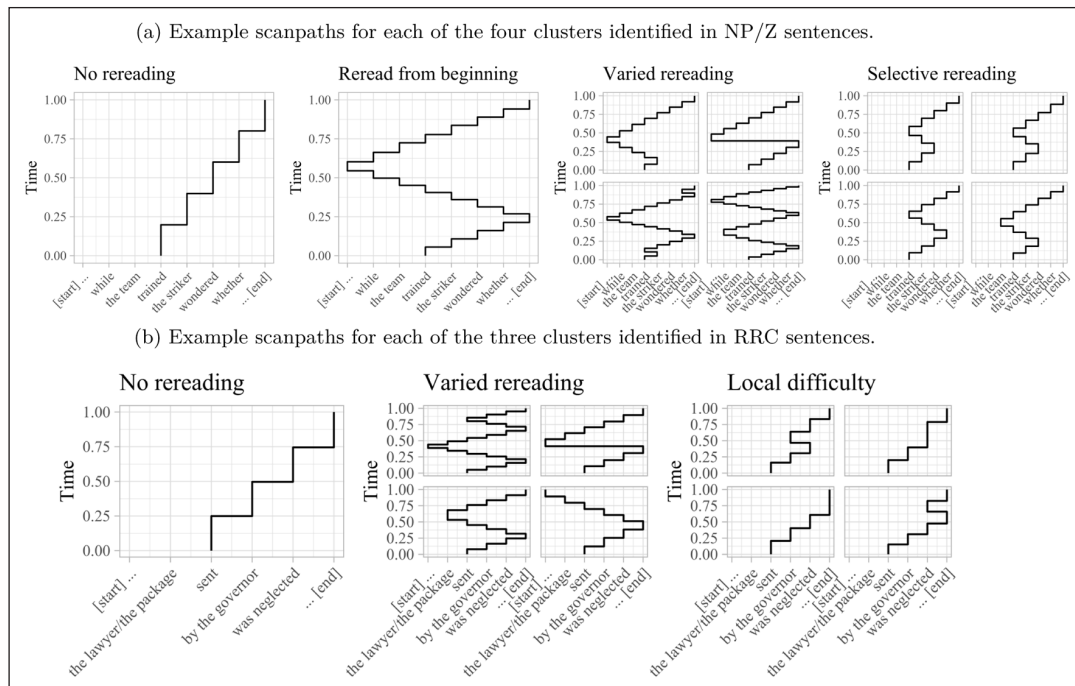


Figure 6: Example scanpaths for NP/Z and RRC sentences.

Additional scanpath clusters were unique to particular sentence types. For NP/Z sentences, a small “selective rereading” cluster emerged: In this cluster, participants returned from the disambiguating region to the ambiguous verb *trained* and then resumed forward reading, consistent with the selective reanalysis hypothesis. By contrast, RRC sentences showed a “local difficulty” cluster in which participants spent a lot of time in one of the post-disambiguation regions and/or made single-region regressions after the point of disambiguation. Finally, reflexive sentences showed two additional clusters: one in which participants regressed from the spillover region to the reflexive region (“reread reflexive”), and one in which participants engaged in more varied but localized rereading after encountering the reflexive (“local rereading”).

Across all sentence types, there were correlations between the rereading patterns and the end-of-trial acceptability judgments. For NP/Z sentences, the “selective reading” and “varied rereading” patterns were accompanied by a steep drop in acceptability of -18% (“selective rereading”) and -20% (“varied rereading”) in the no-comma condition, compared to no rereading. For RRC sentences, “varied rereading” also negatively affected acceptability judgments compared to no rereading (-10%) but “local difficulty” did not. The acceptability of discourse inconsistency sentences generally decreased when rereading was present, across the consistent and inconsistent conditions. Finally, the acceptability of reflexive sentences suffered most noticeably in the “local rereading” cluster compared to no rereading, across the distractor match and mismatch conditions ($-25\%/-17\%$).

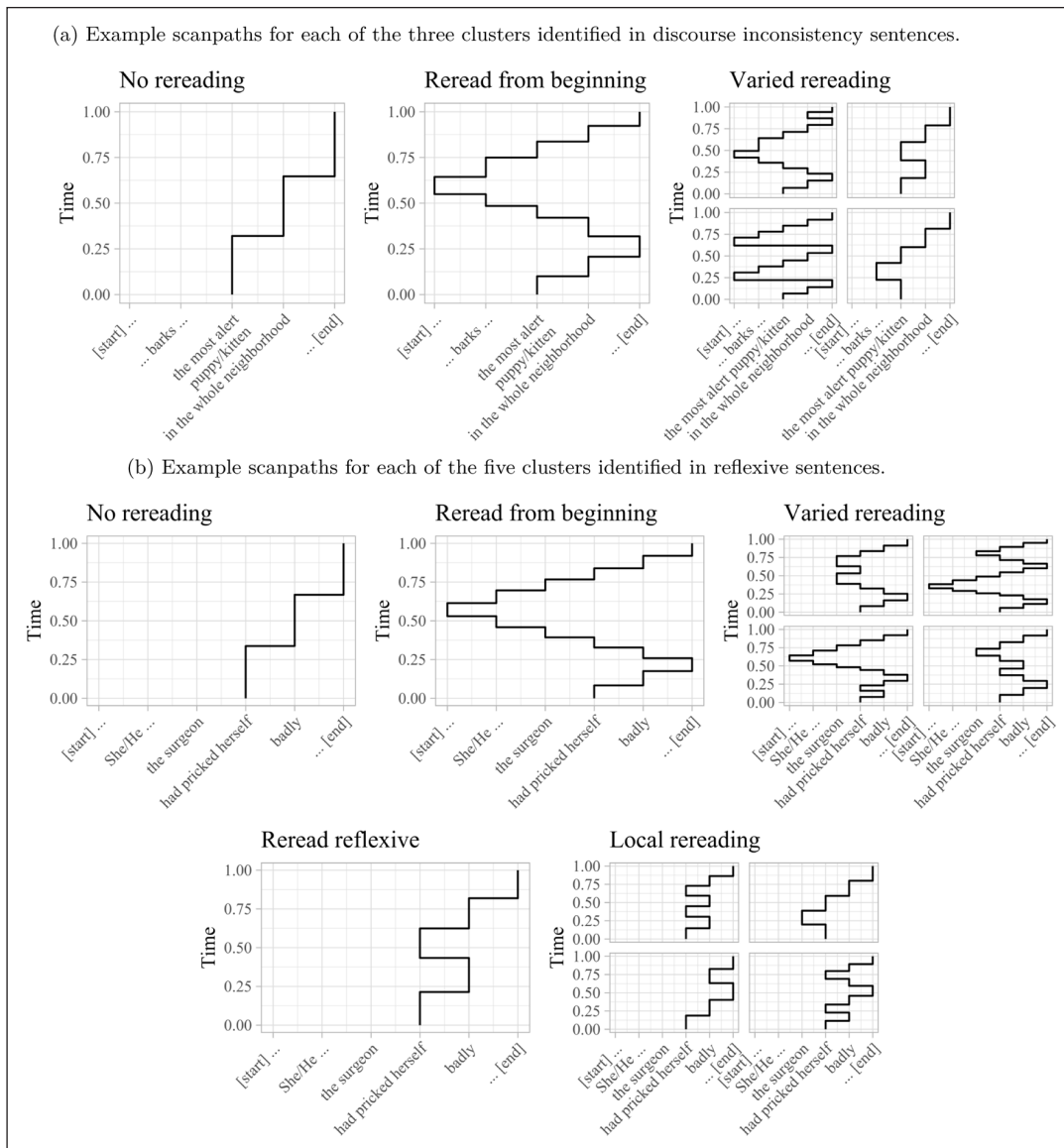


Figure 7: Example scanpaths for discourse inconsistency and reflexive sentences.

11.2 Discussion of scanpath results

The results of the scanpath analyses show that, across all sentence types, participants did not reread at all in a majority of trials. This may indicate that regressions are a kind of last resort in BSPR, and that readers prefer to deal with processing difficulties “covertly”, that is, by pausing at the current position and revisiting earlier regions in memory (Lewis, 1998), rather than overtly. Nevertheless, there was a clear trend across three of the four constructions towards more extensive rereading in the more difficult of the two conditions (NP/Z without comma, RRC with animate NP, inconsistent discourse continuation, reflexives with matching distractor).

A strategy that appeared in three out of the four sentence types was to read until the end of the sentence after visiting the critical region, then move back to the beginning of the sentence in a region-by-region fashion, and finally make a second pass over the entire sentence. Rereading from the beginning of the sentence was also observed in previous studies, both in eye tracking during reading (von der Malsburg and Vasishth, 2011, 2013; Frazier and Rayner, 1982) and in BSPR (Paape and Vasishth, 2022). However, the use of this strategy was largely unaffected by the experimental manipulations in our study, consistent with recent eye-tracking results by Christianson et al. (in prep.).

The scanpath results also reveal that selective rereading occurred in garden-path sentences of the NP/Z type, though only in a small subset of trials and only for about one-third of participants. By contrast, there was no indication of selective rereading in any of the other constructions, apart from very local regressions in RRC and reflexive sentences, which may be confirmatory “checks” (von der Malsburg and Vasishth, 2011, 2013).

A logistic regression revealed that selective rereading of NP/Z sentences was more likely in the no-comma than in the comma condition ($\hat{\Delta} = 1\%$, CrI: [0%, 2%]), consistent with selective reanalysis. However, the pattern was associated with fewer rather than with more positive acceptability judgments, compared to trials with no rereading. Indeed, the drop in acceptability between comma and no-comma sentences was largest in the selective rereading cluster (79% vs 25% acceptable). This observation is more in line with confirmatory rereading than with revisionary rereading as proposed by Frazier and Rayner (1982): If the purpose of selective rereading in NP/Z sentences is to revise the initial analysis, and if we assume that positive judgments are given when the correct structure has been built, acceptability should have increased rather than decreased in the selective rereading cluster. By contrast, if rereading primarily serves to check but not necessarily to revise the first-pass parse (Christianson et al., 2017, in prep.; Levy, 2008; Levy et al., 2009; Bicknell and Levy, 2011), acceptability should not be affected or even decrease, as readers will become more certain that the sentence is ungrammatical. Confirmatory rereading would also explain why the selective rereading cluster contained trials from the unambiguous NP/Z condition, assuming that readers occasionally check their parses even when there is no garden path (see also von der Malsburg and Vasishth 2011).

Among the two remaining sentence types, reflexive sentences showed a greater variety of scanpath clusters compared to discourse inconsistency sentences. This may signal that readers’ reactions to the distractor manipulation in reflexive sentences are more varied than their reactions to the discourse manipulation. In this context, recall that the results of previous studies on the effect of similarity-based interference in reflexive dependencies are inconsistent (Jäger et al., 2017; Parker and Phillips, 2017; Jäger et al., 2020; Dillon et al., 2013). If readers indeed differ widely in their response to this type of interference, this may explain why the effects vary across studies, as certain strategies may be represented in different participant samples to different extents (Yadav et al., 2021). We will now discuss the results’ broader implications.

12. General discussion

The aim of this study was to investigate conscious rereading across different sentence types with the bidirectional self-paced reading (BSPR) paradigm. We were especially interested in distinguishing confirmatory rereading, which serves to increase the reader's certainty in their existing interpretation, from revisionary rereading, which serves to change the existing interpretation. We tested four constructions: NP/Z garden-path sentences, RRC garden-path sentences, sentences with consistent or inconsistent discourse continuations, and sentences with reflexive anaphors that contain a structurally unavailable distractor antecedent.

All our manipulations were effective in the sense that they reliably affected end-of-sentence acceptability judgments. In addition, right-bounded reading times in both of our BSPR experiments showed effects that are consistent with the existing eye-tracking literature: Ambiguous NP/Z sentences showed a strong garden-path effect at the point of disambiguation compared to unambiguous control sentences. RRC garden-path sentences were read more slowly at the point of disambiguation when the initial misanalysis was semantically plausible rather than implausible. Sentences with inconsistent discourse continuations showed a slowdown over several regions compared to sentences with consistent continuations. In reflexive sentences, which show conflicting results in the eye-tracking literature, the only consistent finding was that readers spent more time in the sentence-final region when the sentence contained a feature-matching distractor as opposed to a feature-mismatching distractor.

Despite evidence of processing difficulty across all sentence types, we found consistent evidence for selective rereading only in NP/Z garden-path sentences. The pattern of regressive rereading times in these sentences is compatible with the prediction of Frazier and Rayner (1982) that readers should focus mainly on the ambiguous portion of the string when first-pass parsing fails and syntactic reanalysis is triggered. The scanpath analysis confirmed that there were trials in which readers regressed from the point of disambiguation to the ambiguous region but no further, and then returned to progressive reading. Together, these findings provide the strongest direct evidence for selective rereading in garden-path sentences to date that we are aware of. The original results of Frazier and Rayner (1982) were descriptive in nature, and the observed regressions may have been due to readers trying to “buy time” as opposed to selectively targeting a specific sentence region (Mitchell et al., 2008). Mitchell et al. (2008) found some evidence of targeted rereading in NP/Z sentences, but the associated scanpaths were highly varied.⁹ Similarly, Meseguer et al. (2002) reported evidence of targeted rereading in Spanish garden-path sentences, but von der Malsburg and Vasishth (2011) could not identify a unique, selective scanpath pattern in Meseguer et al.'s data.

⁹ Green (2013) conducted scanpath analyses of the Mitchell et al. (2008) eye-tracking data and did find scanpath clusters consistent with selective rereading, but there was no evidence that the proportions of these clusters were affected by the garden-path manipulation.

However, our results as a whole do not support the selective reanalysis hypothesis: There was no indication in the data that selective rereading helped readers find a grammatical analysis of the sentence, as evidenced by the large proportion of negative acceptability judgments following selective rereading. Thus, if a selective reanalysis mechanism exists, it would have to have a very high failure rate, which appears inconsistent with the claim by Frazier and Rayner (1982) that it is the parser's default mechanism for dealing with temporary ambiguity. As noted by Fodor and Inoue (1994), the selective reanalysis hypothesis focuses on the diagnosis portion of garden-path recovery: The main challenge is to locate the source of the processing difficulty, and once the parser has succeeded, recovery is assumed to be straightforward. Thus, the presence of selective rereading in a given trial should indicate that diagnosis, and therefore recovery, was successful, contrary to what our data show.

Our result is in line with the findings of Christianson et al. (2017, in prep.), who found no evidence in their eye-tracking experiments that selective rereading is a frequent strategy, or that it leads to more successful comprehension of NP/Z sentences. The scanpath pattern associated with selective rereading was also very rare in our BSPR data: 4% of trials showed the selective strategy, and only about 30% of participants ever made use of it. The largest group of scanpaths that included rereading, encompassing the “varied rereading” and “reread from beginning” clusters, most frequently showed some variant of rereading the entire sentence.

Considering the absence of a positive effect of selective rereading on sentence acceptability, our results are more in line with a confirmatory view of selective rereading than with a revisionary view (Christianson et al., 2017, in prep.; Levy, 2008; Levy et al., 2009; Bicknell and Levy, 2011). Conscious, selective rereading at the sentence level may be limited to the NP/Z garden-path structure and possibly other extremely difficult structures. We found no consistent indication of selective rereading in RRC garden-path sentences, which are milder than NP/Z garden-path sentences (Van Schijndel and Linzen, 2021). The BSPR experiments of Paape and Vasishth (2022) showed no indication of selective rereading in German garden-path sentences, which are presumably also milder than NP/Z sentences. By contrast, studies that did find evidence of selective rereading are mostly limited to the NP/Z construction (Frazier and Rayner, 1982; Mitchell et al., 2008; Schotter et al., 2014), with the exception of Meseguer et al. (2002), who investigated adverbial clauses with ambiguous attachment in Spanish. This may suggest that there are two classes of garden paths: those that can be covertly or “automatically” reanalyzed and those that cannot. The underlying reason for such a distinction could be that the parser is missing the necessary reanalysis mechanisms for some structures (e.g., Marcus, 1980; Pritchett, 1992; Lewis, 1998) or that some structures involve early “pruning” of the correct alternative analysis, which thus cannot be recovered without conscious effort (see discussion in Meng and Bader, 2000). The NP/Z construction would appear to belong to this class of structures.

Another interesting property of the NP/Z structure became apparent from participants' comments during debriefing. Ten participants commented that commas had been missing from some of the

sentences, presumably referring to NP/Z sentences. Comma-less NP/Z sentences in combination with acceptability judgments have been used in a number of psycholinguistic studies (Van Dyke and Lewis, 2003; Tabor and Hutchins, 2004; Ferreira and Henderson, 1991; McElree, 1993). However, a check of several online grammar resources (Grammarly blog, 2021; [Grammar.com](https://www.grammar.com), 2021; Purdue OWL, 2021) reveals that the use of a comma after an “introductory clause” (*While the team trained, the striker wondered ...*) is considered mandatory, though exceptions may be made for “short” introductory phrases. In our BSPR sample, 13 participants (out of 193) always judged comma-less NP/Z sentences as unacceptable. Descriptively, these participants were more likely to use the selective rereading strategy in NP/Z sentences than the rest of the sample (10% versus 4% of trials). It is thus possible that selective rereading was partly due to participants checking whether a comma was present, and rejecting the sentence as ungrammatical when it was missing (see also Staub, 2007). However, even if all selective regressions in our data were triggered by the missing comma, our conclusion would not change: Conscious rereading appears to be due to the need to verify previously-read information, rather than by the need to reanalyze the syntactic structure of the sentence.

In addition, participants may have penalized sentences that were difficult to process and required rereading (Hofmeister et al., 2013), even though our instructions were to judge the sentences based on their grammaticality and meaning. Such a rereading penalty may partly explain the negative effect of rereading on acceptability that we observed across all sentence types. However, this account would still not favor the selective reanalysis hypothesis: If selective rereading is the parser’s default reanalysis mechanism, as claimed by Frazier and Rayner (1982), because it is arguably more efficient than rereading the entire sentence, it would be surprising if selective reanalysis incurred so much additional processing difficulty that participants end up rejecting the sentence. Alternatively, it could be that the NP/Z structure in particular is so difficult to reanalyze that it leads to rejections even when participants ultimately succeed at reanalysis. However, it was precisely the NP/Z structure that Frazier and Rayner used to argue for selective reanalysis, which makes a purely effort-based explanation of the negative judgments less convincing.

It might also be argued that any rereading observed in our experiments is an artifact of the judgment task, and that participants may have regressed even less if we had used comprehension questions or no task at all. Having participants read sentences without any task would be the only way to investigate reading uncontaminated by task effects. The downside is that one would lose all information about the final processing outcome. The usual way to deal with this issue is to use comprehension questions, which, however, can also affect reading strategies (Swets et al., 2008). Furthermore, it is difficult to establish a direct link between participants’ answers to comprehension questions and their syntactic analysis of a given sentence, as the responses may be affected by additional inferential and memory processes (Huang and Ferreira, 2021; Bader and Meng, 2018; Meng and Bader, 2021; Qian et al., 2018). Because our goal was to investigate

the reading process along with the final processing outcome, we needed to add an offline task, in the same way as previous studies on selective reanalysis (Schotter et al., 2014; Meseguer et al., 2002; Frazier and Rayner, 1982). The linking assumption behind the judgment task was that participants give a negative judgment if they are unable to find a grammatical analysis of the string (Ferreira and Henderson, 1991; Warner and Glass, 1987). Investigating the effect of different tasks such as comprehension questions, acceptability ratings etc. on rereading patterns seems like an important future direction for psycholinguistics, but is beyond the scope of the current project.

Regarding the role of conscious awareness in rereading, it is tempting to conclude that the likelihood of deliberate, confirmatory rereading at the sentence level depends on the subtlety of the linguistic manipulation as reflected in the offline measures: Trials in which readers decide to reread parts of the sentence may be a subset of the trials in which readers decide that a sentence is malformed and ultimately reject it. In line with this view, the low offline acceptability of comma-less NP/Z sentences coincided with partial evidence of selective rereading during online processing. By contrast, RRC sentences with animate NPs and reflexive sentences with matching distractors showed relatively high acceptability and no evidence of selective rereading. However, the apparent connection between negative judgments and selective rereading is broken by sentences with inconsistent discourse continuations: As for NP/Z sentences, acceptability was low for this construction, yet there was no evidence of selective rereading. A plausible explanation for this mismatch is that difficulties that arise at the level of world knowledge (“Kittens do not bark”) are distinct from difficulties at the level of the syntactic analysis: Selective rereading may be more likely as a response to conscious parsing failure, because participants want to make sure that there is no correct alternative parse. By contrast, when the discourse is inconsistent (*barks ... kitten*), it makes little sense to explore alternative analyses, given that the sentence is already well-formed from a purely syntactic point of view. Furthermore, it is unlikely that an adult native speaker would misread *meows* in the first sentence as *barks*, or *puppy* in the second sentence as *kitten*. Confirmatory rereading is thus not needed, but the sentence is still confidently rejected.

This perspective is in line with the proposal that readers are sensitive to the relative probabilities of different types of linguistic errors in their environment, such as misspelled, missing or superfluous words (Ryskin et al., 2018, 2021; Gibson et al., 2013). In addition, readers may be sensitive to the types of *processing errors* that they could have plausibly made while reading the sentence. Parsing errors appear to be intuitively judged as likely, and are thus more likely to trigger conscious, confirmatory rereading. By contrast, gross errors in word identification may be judged as relatively unlikely, and are thus less likely to trigger confirmatory rereading, at least in adults. Future work could investigate this possibility by collecting data about individual participants’ perceived relative probabilities of different processing errors, and correlating these with the corresponding online processing profiles.

A final salient question about our results is whether they are in line with the existing eye-tracking literature. The BSPR paradigm differs from eye tracking in that reading is more deliberate and less automatic. However, despite the clear differences between the paradigms, it is still informative to see whether the methods yield converging evidence for one and the same phenomenon. We focus here on first-pass regressions from the disambiguating region in NP/Z garden-path sentences (*While the team trained(,) the striker wondered ...*). **Figure 8** shows the estimated effect from our combined BSPR data from both experiments (193 subjects) compared to estimates from five published eye-tracking studies that report regression proportions in NP/Z sentences.

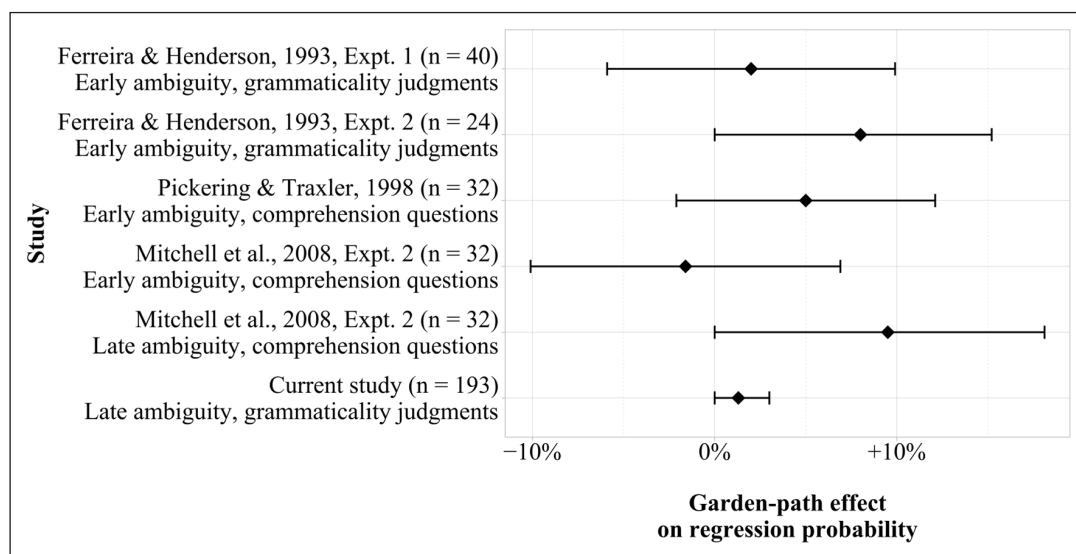


Figure 8: Comparison of garden-path effects in first-pass regression probabilities between five published eye-tracking studies and the current BSPR study. Error bars show 95% confidence intervals.

As the figure shows, the estimates from the eye-tracking studies of Ferreira and Henderson (1993), Pickering and Traxler (1998), and Mitchell et al. (2008) all have much larger uncertainty intervals than our BSPR study; these large uncertainties are due to the smaller sample sizes used in these studies (24–40 participants, cf. our 193 participants). Some of the experiments differed from ours not only in terms of the experimental method, but also with regard to the end-of-trial task (grammaticality judgments versus comprehension questions), and/or with regard to where in the sentence the ambiguity appeared (early versus late). Nevertheless, our estimate of a 0–3% increase in regressions at the disambiguating region is consistent with the overall empirical picture: The BSPR uncertainty interval is entirely contained in the uncertainty intervals from the earlier eye-tracking studies.

Despite the observed overlap between our BSPR-based estimate and those from the eye-tracking studies, it should be noted that the eye-tracking studies have a much higher baseline

of regressions in unambiguous (comma) condition. For instance, in their Experiment 2, Mitchell et al. (2008) observed a regression rate of 18% at the critical verb *wondered* even in the comma condition (*While the team trained, the striker wondered ...*). By contrast, in our BSPR study, the mean regression rate at this position in the comma condition was 2%. Compared to eye tracking, BSPR eliminates several potential sources of regressions, such as overshoot correction (Rayner, 1998) or “time-out” regressions that allow higher-level cognitive processes to catch up to the forward movement of the eyes (Mitchell et al., 2008; Engelmann et al., 2013). Overshoot correction in particular may be the source of many regressions in eye tracking: For instance, Eskenazi and Folk (2017) report a regression rate of 17% following an infrequent three-letter word (... wore a green gem around her neck ...), 89% of which were classified as corrective rather than comprehension-mediated. Given that corrective regressions are likely automatic, it is thus not surprising that BSPR, which only allows deliberate regressions, has a much lower baseline regression probability. In future work, we aim to design experiments that specifically probe for conscious deliberation in eye-tracking regressions, which may uncover shared mechanisms between deliberate regressions in BSPR and in eye tracking.

13. Conclusion

Using an experimental paradigm in which regressions during rereading are carried out consciously, our study showed that selective rereading occurs only as a consequence of a difficult syntactic misparse (garden path), and that selective rereading, when carried out consciously, is confirmatory: The reader revisits earlier material selectively to confirm their initial analysis or to confirm the perceived ungrammaticality of the sentence, rather than to revise an initial misanalysis.

Data availability statement

All experimental materials, data and analysis code are available at:

<https://doi.org/10.17605/OSF.IO/54BDA>.

Funding information

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant 428960187 – PI: Dario Paape.

Acknowledgements

The authors would like to thank Titus von der Malsburg, Kiel Christianson, the members of the Vasishth Lab, and the audiences at CUNY 2021 and AMLaP 2021 for helpful comments and suggestions.

Competing interests

The authors have no competing interests to declare.

Author contributions

Dario Paape: Funding Acquisition, Conceptualization, Methodology, Software, Investigation, Formal Analysis, Writing – Original Draft, Project Administration; Shravan Vasishth: Funding Acquisition, Conceptualization, Writing – Review & Editing, Supervision.

References

- Aaronson, D., & Scarborough, H. S. (1976). Performance theories for sentence coding: Some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 56–70. DOI: [https://doi.org/10.1016/S0022-5371\(77\)80052-2](https://doi.org/10.1016/S0022-5371(77)80052-2)
- Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2015). A comparison of scanpath comparison methods. *Behavior Research Methods*, 47, 1377–1392. DOI: <https://doi.org/10.3758/s13428-014-0550-3>
- Bader, M., & Meng, M. (2018). The misinterpretation of noncanonical sentences revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44, 1286–1311. DOI: <https://doi.org/10.1037/xlm0000519>
- Bicknell, K., & Levy, R. (2011). Why readers regress to previous words: A statistical analysis. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. URL: <https://escholarship.org/uc/item/7jf4w8sv>.
- Booth, R. W., & Weger, U. W. (2013). The function of regressions in reading: Backward eye movements allow rereading. *Memory & Cognition*, 41, 82–97. DOI: <https://doi.org/10.3758/s13421-012-0244-y>

- Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. DOI: <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P. C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*, 395–411. DOI: <https://doi.org/10.32614/RJ-2018-017>
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Christianson, K., Dempsey, J., Tsiola, A., & Goldshtein, M. (2022). What if they're just not that into you (or your experiment)? On motivation and psycholinguistics. *Psychology of Learning and Motivation*, *76*, 51–88. DOI: <https://doi.org/10.1016/bs.plm.2022.03.002>
- Christianson, K., Luke, S. G., Hussey, E. K., & Wochna, K. L. (2017). Why reread? Evidence from garden-path and local coherence structures. *Quarterly Journal of Experimental Psychology*, *70*, 1380–1405. DOI: <https://doi.org/10.1080/17470218.2016.1186200>
- Christianson, K., Tsiola, A., Deshaies, S. E., & Kim, N. (in prep). *Nonselective rereading of garden-path sentences: Evidence from reading times, comprehension, and scanpaths*. Manuscript in preparation.
- Clifton, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*, 317–334. DOI: [https://doi.org/10.1016/S0749-596X\(03\)00070-6](https://doi.org/10.1016/S0749-596X(03)00070-6)
- Connor, C. M., Radach, R., Vorstius, C., Day, S. L., McLean, L., & Morrison, F. J. (2015). Individual differences in fifth graders' literacy and academic language predict comprehension monitoring development: An eye-movement study. *Scientific Studies of Reading*, *19*, 114–134. DOI: <https://doi.org/10.1080/10888438.2014.943905>
- Cromley, J. G., & Azevedo, R. (2006). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning*, *1*, 229–247. DOI: <https://doi.org/10.1007/s11409-006-9002-5>
- Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*, 85–103. DOI: <https://doi.org/10.1016/j.jml.2013.04.003>
- Drummond, A. (2013). Ibx farm. Online server: <http://spellout.net/ibexfarm>
- Eilers, S., Tiffin-Richards, S. P., & Schroeder, S. (2018). Individual differences in children's pronoun processing during reading: Detection of incongruence is associated with higher reading fluency and more regressions. *Journal of Experimental Child Psychology*, *173*, 250–267. DOI: <https://doi.org/10.1016/j.jecp.2018.04.005>
- Engelmann, F., Vasishth, S., Engbert, R., & Kliegl, R. (2013). A framework for modeling the interaction of syntactic processing and eye movement control. *Topics in Cognitive Science*, *5*, 452–474. DOI: <https://doi.org/10.1111/tops.12026>
- Eskenazi, M. A., & Folk, J. R. (2017). Regressions during reading: The cost depends on the cause. *Psychonomic Bulletin & Review*, *24*, 1211–1216. DOI: <https://doi.org/10.3758/s13423-016-1200-9>
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368. DOI: [https://doi.org/10.1016/0749-596X\(86\)90006-9](https://doi.org/10.1016/0749-596X(86)90006-9)

- Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30, 725–745. DOI: [https://doi.org/10.1016/0749-596X\(91\)90034-H](https://doi.org/10.1016/0749-596X(91)90034-H)
- Ferreira, F., & Henderson, J. M. (1993). Reading processes during syntactic analysis and reanalysis. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 47, 247–275. DOI: <https://doi.org/10.1037/h0078819>
- Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research*, 23, 407–434. DOI: <https://doi.org/10.1007/BF02143947>
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178–210. DOI: [https://doi.org/10.1016/0010-0285\(82\)90008-1](https://doi.org/10.1016/0010-0285(82)90008-1)
- Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Ph.D. thesis. Carnegie Mellon University.
- Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, 110, 8051–8056. DOI: <https://doi.org/10.1073/pnas.1216438110>
- Godfroid, A., Loewen, S., Jung, S., Park, J. H., Gass, S., & Ellis, R. (2015a). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition*, 37, 269–297. DOI: <https://doi.org/10.1017/S0272263114000850>
- Godfroid, A., Winke, P., & Rebuschat, P. (2015b). Investigating implicit and explicit processing using L2 learners' eye-movement data. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (Vol. 48, p. 325). Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/sibil.48.14god>
- Gong, T. (2019). FAB: A dummy's program for self-paced forward and backward reading. URL: <https://github.com/tianweigong/fabreading>
- Grammar.com: Commas and introductory clauses or phrases*. (2021). <https://www.grammar.com/commas-and-introductory-clauses-or-phrases/>. Accessed: 2021-08-22.
- Grammarly blog: Commas after introductory clauses*. (2021). <https://www.grammarly.com/blog/commas-after-introductory-clauses/>. Accessed: 2021-08-22.
- Green, M. J. (2013). *On Repairing Sentences: An Experimental and Computational Analysis of Recovery from Unexpected Syntactic Disambiguation in Sentence Parsing*. Ph.D. thesis. University of Exeter.
- Hatfield, H. (2016). Self-guided reading: Touch-based measures of syntactic processing. *Journal of Psycholinguistic Research*, 45, 121–141. DOI: <https://doi.org/10.1007/s10936-014-9334-2>
- Hessel, A. K., Nation, K., & Murphy, V. A. (2021). Comprehension monitoring during reading: An eye-tracking study with children learning English as an additional language. *Scientific Studies of Reading*, 25, 159–178. DOI: <https://doi.org/10.1080/10888438.2020.1740227>

- Hofmeister, P., Jaeger, T. F., Arnon, I., Sag, I. A., & Snider, N. (2013). The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes*, 28, 48–87. DOI: <https://doi.org/10.1080/01690965.2011.572401>
- Huang, Y., & Ferreira, F. (2021). What causes lingering misinterpretations of garden-path sentences: Incorrect syntactic representations or fallible memory processes? *Journal of Memory and Language*, 121, 104288. DOI: <https://doi.org/10.1016/j.jml.2021.104288>
- Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, 94, 316–339. DOI: <https://doi.org/10.1016/j.jml.2017.01.004>
- Jäger, L. A., Merten, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*, 111, 104063. DOI: <https://doi.org/10.1016/j.jml.2019.104063>
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354. DOI: <https://doi.org/10.1037/0033-295X.87.4.329>
- Kemtes, K. A., Brennan, A., & Wingfield, A. (2001). *Allocation and reallocation of sentence processing time: A view from cognitive aging*. Poster presented at the 42nd Annual Meeting of the Psychonomic Society. DOI: <https://doi.org/10.1037/e537102012-341>
- Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29, 1–27. DOI: <https://doi.org/10.1007/BF02289565>
- Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29, 115–129. DOI: <https://doi.org/10.1007/BF02289694>
- Leininger, M., Myslín, M., Rayner, K., & Levy, R. (2017). Do resource constraints affect lexical processing? Evidence from eye movements. *Journal of Memory and Language*, 93, 82–103. DOI: <https://doi.org/10.1016/j.jml.2016.09.002>
- Levy, R. (2008). A noisy-channel model of human sentence comprehension under uncertain input. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 234–243. DOI: <https://doi.org/10.3115/1613715.1613749>
- Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, 106, 21086–21090. DOI: <https://doi.org/10.1073/pnas.0907664106>
- Lewis, R. L. (1998). Reanalysis and limited repair parsing: Leaping off the garden path. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 247–285). Dordrecht: Kluwer. DOI: https://doi.org/10.1007/978-94-015-9070-9_8
- Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 13, 123–146. DOI: <https://doi.org/10.1080/1057356970130203>
- Marcus, M. P. (1980). *A theory of syntactic recognition for natural language*. Cambridge, MA: MIT Press.
- McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory and Language*, 32, 536–571. DOI: <https://doi.org/10.1006/jmla.1993.1028>

- Meng, M., & Bader, M. (2000). Mode of disambiguation and garden-path strength: An investigation of subject-object ambiguities in German. *Language and Speech*, *43*, 43–74. DOI: <https://doi.org/10.1177/00238309000430010201>
- Meng, M., & Bader, M. (2021). Does comprehension (sometimes) go wrong for noncanonical sentences? *Quarterly Journal of Experimental Psychology*, *74*, 1–28. DOI: <https://doi.org/10.1177/1747021820947940>
- Meseguer, E., Carreiras, M., & Clifton, C. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, *30*, 551–561. DOI: <https://doi.org/10.3758/BF03194956>
- Metzner, P., von der Malsburg, T., Vasishth, S., & Rösler, F. (2017). The importance of reading naturally: Evidence from combined recordings of eye movements and electric brain potentials. *Cognitive Science*, *41*, 1232–1263. DOI: <https://doi.org/10.1111/cogs.12384>
- Mitchell, D. C., & Green, D. W. (1978). The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology*, *30*, 609–636. DOI: <https://doi.org/10.1080/14640747808400689>
- Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the Selective Reanalysis hypothesis. *Journal of Memory and Language*, *59*, 266–293. DOI: <https://doi.org/10.1016/j.jml.2008.06.002>
- Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, *29*, 186–217. DOI: <https://doi.org/10.1080/01690965.2012.758297>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2020). vegan: Community Ecology Package. URL: <https://CRAN.R-project.org/package=vegan>. r package version 2.5-7.
- Paape, D., & Vasishth, S. (2022). Is reanalysis selective when regressions are consciously controlled? *Glossa Psycholinguistics*, *1*. DOI: <https://doi.org/10.5070/G601139>
- Paape, D., Vasishth, S., & Engbert, R. (2021). Does local coherence lead to targeted regressions and illusions of grammaticality? *Open Mind*, 42–58. DOI: https://doi.org/10.1162/opmi_a_00041
- Palan, S., & Schitter, C. (2018). Prolific.ac – A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. DOI: <https://doi.org/10.1016/j.jbef.2017.12.004>
- Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language*, *94*, 272–290. DOI: <https://doi.org/10.1016/j.jml.2017.01.002>
- Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 940–961. DOI: <https://doi.org/10.1037/0278-7393.24.4.940>
- Pritchett, B. L. (1992). Grammatical competence and parsing performance. University of Chicago Press, Chicago, IL.
- Purdue, O. W. L. (2021). Commas after introductions. https://owl.purdue.edu/owl/general_writing/punctuation/commas/commas_after_introductions.html. Accessed: 2021-08-22.

- Qian, Z., Garnsey, S., & Christianson, K. (2018). A comparison of online and offline measures of good-enough processing in garden-path sentences. *Language, Cognition and Neuroscience*, 33, 227–254. DOI: <https://doi.org/10.1080/23273798.2017.1379606>
- R Core Team. (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422. DOI: <https://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358–374. DOI: [https://doi.org/10.1016/S0022-5371\(83\)90236-0](https://doi.org/10.1016/S0022-5371(83)90236-0)
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, 10, 241–255. DOI: https://doi.org/10.1207/s1532799xssr1003_3
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105, 125–157. DOI: <https://doi.org/10.1037/0033-295X.105.1.125>
- Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition*, 181, 141–150. DOI: <https://doi.org/10.1016/j.cognition.2018.08.018>
- Ryskin, R., Stearns, L., Bergen, L., Eddy, M., Fedorenko, E., & Gibson, E. (2021). An erp index of realtime error correction within a noisy-channel framework of human communication. *Neuropsychologia*, 158, 107855. DOI: <https://doi.org/10.1016/j.neuropsychologia.2021.107855>
- Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): Comprehension is supported by regressions during reading. *Psychological Science*, 25, 1218–1226. DOI: <https://doi.org/10.1177/0956797614531148>
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8, 289–317. DOI: <https://doi.org/10.32614/RJ-2016-021>
- Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 550–569. DOI: <https://doi.org/10.1037/0278-7393.33.3.550>
- Staub, A. (2011). Word recognition and syntactic attachment in reading: Evidence for a staged architecture. *Journal of Experimental Psychology: General*, 140, 407–433. DOI: <https://doi.org/10.1037/a0023517>
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542–562. DOI: [https://doi.org/10.1016/S0749-596X\(02\)00536-3](https://doi.org/10.1016/S0749-596X(02)00536-3)
- Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*, 40, 136–150. DOI: <https://doi.org/10.1006/jmla.1998.2606>

- Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*, 201–216. DOI: <https://doi.org/10.3758/MC.36.1.201>
- Tabor, W., & Hutchins, S. (2004). Evidence for self-organized sentence processing: digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 431–450. DOI: <https://doi.org/10.1037/0278-7393.30.2.431>
- 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*, 285–318. DOI: <https://doi.org/10.1006/jmla.1994.1014>
- Van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, *54*, 360–372. DOI: <https://doi.org/10.1080/0163853X.2017.1306677>
- Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*, 285–316. DOI: [https://doi.org/10.1016/S0749-596X\(03\)00081-0](https://doi.org/10.1016/S0749-596X(03)00081-0)
- van Moort, M. L., Koornneef, A., & van den Broek, P. W. (2018). Validation: Knowledge-and text-based monitoring during reading. *Discourse Processes*, *55*, 480–496. DOI: <https://doi.org/10.1080/0163853X.2018.1426319>
- Van Schijndel, M., & Linzen, T. (2021). Single-stage prediction models do not explain the magnitude of syntactic disambiguation difficulty. *Cognitive Science*, *45*, e12988. DOI: <https://doi.org/10.1111/cogs.12988>
- Vasishth, S., & Drenhaus, H. (2011). Locality in German. *Dialogue & Discourse*, *2*, 59–82. DOI: <https://doi.org/10.5087/dad.2011.104>
- Vasishth, S., & Gelman, A. (2021). How to embrace variation and accept uncertainty in linguistic and psycholinguistic data analysis. *Linguistics*, *59*, 1311–1342. DOI: <https://doi.org/10.1515/ling-2019-0051>
- Vidal-Abarca, E., Martínez, T., Salmerón, L., Cerdán, R., Gilabert, R., Gil, L., Mañá, A., Llorens, A. C., & Ferris, R. (2011). Recording online processes in task-oriented reading with Read&Answer. *Behavior Research Methods*, *43*, 179–192. DOI: <https://doi.org/10.3758/s13428-010-0032-1>
- von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, *94*, 119–133. DOI: <https://doi.org/10.1016/j.jml.2016.10.003>
- von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, *65*, 109–127. DOI: <https://doi.org/10.1016/j.jml.2011.02.004>
- von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*, 1545–1578. DOI: <https://doi.org/10.1080/01690965.2012.728232>
- Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language*, *26*, 714–738. DOI: [https://doi.org/10.1016/0749-596X\(87\)90111-2](https://doi.org/10.1016/0749-596X(87)90111-2)

- Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z Reader. *Cognition*, *111*, 132–137. DOI: <https://doi.org/10.1016/j.cognition.2008.12.011>
- Weiss, A. F. (2020). The information gathering framework – a cognitive model of regressive eye movements during reading. *Journal of Eye Movement Research*, *13*. DOI: <https://doi.org/10.16910/jemr.13.4.4>
- Yadav, H., Paape, D., Smith, G., Dillon, B., & Vasishth, S. (2021). Individual differences in cueweighting in sentence comprehension: An evaluation using Approximate Bayesian Computation. URL: psyarxiv.com/4jdu5. DOI: <https://doi.org/10.31234/osf.io/4jdu5>
- Yan, S., & Jaeger, T. F. (2020). Expectation adaptation during natural reading. *Language, Cognition and Neuroscience*, *35*, 1394–1422. DOI: <https://doi.org/10.1080/23273798.2020.1784447>
- Zargar, E., Adams, A. M., & Connor, C. M. (2020). The relations between children’s comprehension monitoring and their reading comprehension and vocabulary knowledge: An eye-movement study. *Reading and Writing*, *33*, 511–545. DOI: <https://doi.org/10.1007/s11145-019-09966-3>
- Zhang, X., & Witzel, J. (2021). The interaction of semantic information and parsing biases: An A-maze investigation. *34th Annual CUNY Conference on Human Sentence Processing*.

