

Is reanalysis selective when regressions are consciously controlled?

Dario Paape, University of Potsdam, Germany, paape@uni-potsdam.de

Shravan Vasishth, University of Potsdam, Germany, vasishth@uni-potsdam.de

The selective reanalysis hypothesis of Frazier and Rayner (1982) states that when faced with the need to reanalyze a syntactic ambiguity, readers direct their eyes towards the region in the sentence inducing the ambiguity (e.g., *Since Jay always jogs a mile seems like a short distance to him*). Given the mixed evidence for this proposal in the literature, we investigated the possibility that selective reanalysis is tied to conscious awareness of the garden-path effect. To this end, we adapted the well-known self-paced reading paradigm to allow for regressive as well as progressive key presses. Assuming that regressions in such a paradigm are consciously controlled, we found no evidence for selective reanalysis, but rather for occasional extensive, heterogeneous rereading of garden-path sentences. We discuss the implications of our findings for the selective reanalysis hypothesis, the role of awareness in sentence processing, as well as the usefulness of the bidirectional self-paced reading method for sentence processing research.



1. Introduction

During normal reading, readers do not consciously and deliberately direct their gaze towards each word in the sentence (e.g., Reichle, Pollatsek, Fisher, & Rayner, 1998). In unimpaired, adult native speakers, reading has many attributes of an automatic cognitive process: it is usually fast and effortless, occurs without conscious intention when text is presented, and the components of a reading episode – for instance, having fixated certain words in a sentence while skipping others – cannot be remembered and reported afterwards (Logan, 1997).

Garden-path sentences like in (1) constitute an interesting case with regard to conscious awareness and the automaticity of eye movements in reading:

- (1) a. Mary forgot her husband **needed** a ride yesterday.
 b. Since Jay always jogs a mile and a half **seems** like a very short distance to him.
 c. The boat floated down the river **sank**.

In all of these sentences, the human sentence processor initially prefers to analyze the underlined part as a structural and semantic unit. When the reader reaches the boldfaced word, the syntactic structure of the sentence has to be revised:¹ In (1a,b), *her husband* and *a mile and a half* are not the direct objects of *forgot* and *jogs*, respectively, but rather the subjects of their own clauses, while (1c) contains a reduced relative clause (*The boat [that was] floated ...*). The garden paths in (1) form a continuum with regard to processing difficulty (Sturt, 1996): despite causing measurable slowdowns in reading speed compared to a control condition (Sturt, Pickering, & Crocker, 1999), (1a) subjectively does not result in conscious difficulty. By contrast, (1b) anecdotally often results in a conscious “double take” upon encountering *seems*. This double take is even more noticeable when encountering *sank* in (1c), to the point that the sentence is often rejected as ungrammatical by untrained native speakers (Devitt, 2006). Many theories of parsing have posited a connection between the syntactic operations required for reanalysis and the conscious impression of a garden path (e.g., Lewis, 1998; Marcus, 1980; Pritchett, 1992). Here, we focus on the more readily observable connection between conscious “double takes” and patterns of rereading.

Many eye-tracking studies have found that more regressive saccades and longer refixations occur in garden-path sentences compared to unambiguous control sentences (e.g. Christianson, Tsiola, Deshaies, & Kim, in prep.; Jacob & Felser, 2016; Pickering & Traxler, 1998; Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013). An important linking assumption between syntactic reanalysis and eye movements was proposed by Frazier and Rayner (1982) in the form of the *selective reanalysis hypothesis* (SRH) (p. 182):

¹ This assumes that the parser has already settled on a particular structure before the point of disambiguation, as assumed by most serial parsing models. Parallel models and “minimal commitment” models treat reanalysis effects differently; see Lewis (1998) for discussion.

[T]he parser will use whatever information indicates that its initial analysis is inappropriate to attempt to diagnose the source of its error. If successful, this would permit it to selectively focus on just that portion of the analysis which was responsible for the particular problem it encountered with its first analysis.

The SRH predicts that eye movements during syntactic reanalysis should preferentially target the point of disambiguation and the ambiguous region, as they constitute the symptom and the source of the error, respectively (Fodor & Inoue, 1994). Importantly, selective reanalysis is not conscious: It is assumed that readers are usually unaware of the fact that they have been garden-pathed, as well as of their subsequent recovery strategies. On the contrary, (Frazier & Rayner, 1982, p. 182) assume that conscious awareness of garden-pathing is correlated with a *lack* of selectivity, as it signals that “the parser’s normal correction routines” have failed and the entire sentence needs to be reread.

Using sentences like (1b), Frazier and Rayner (1982) claimed to have found evidence in favor of the SRH: Compared to control sentences, garden-path sentences showed more regressions to the ambiguous noun phrase *a mile and a half*. However, Mitchell, Shen, Green, and Hodgson (2008) later noted that these short-range regressions may not constitute evidence for selectivity: After reaching the point of disambiguation, readers may regress to the immediately preceding region simply to stall reading and buy time for syntactic reanalysis. When Mitchell et al. adapted their materials to eliminate the adjacency confound, only partial, indirect evidence for selectivity was found. Readers did tend to refixate the ambiguous region more often compared to the control condition, but only in a minority of trials, and often only after having previously fixated other regions. Overall, the evidence in favor of selectivity during reanalysis is mixed: Meseguer, Carreiras, and Clifton (2002) report selective rereading for mild garden paths in Spanish, but von der Malsburg and Vasishth’s (2011) analysis of the same data point more towards unselective rereading of the entire sentence, as do the results of von der Malsburg and Vasishth (2013). By contrast, Schotter, Tran, and Rayner (2014) did find some evidence of selective rereading in sentences similar to (1b), in the form of small increase of about 2% in regressions to the beginning of the ambiguous region compared to a control condition. Most recently, Christianson, Luke, Hussey, and Wochna (2017) also reported more mixed evidence for selectivity, as well as no gains in comprehension accuracy when refixations on the critical regions occurred, which casts doubt on the usefulness of the hypothesized revision procedure (see also Christianson et al., in prep.).

What is clear from the literature is that readers do not *always* engage in selective rereading in garden-path sentences. This is obvious even from the original study of Frazier and Rayner (1982), where unselective rereading was observed in a large subset of trials. In other trials, subjects may succeed at reanalyzing the structure “covertly”, without regressing (Lewis, 1998;

von der Malsburg & Vasishth, 2013). However, it is possible that selective rereading occurs stochastically, depending on the properties of the sentences and differences between readers. Under this assumption, it would not be surprising that the empirical picture is not clear, given that individual studies may be insufficiently powered to detect effects that only occur rarely. Furthermore, if selective reanalysis only occurs on a subset of trials, it may not be an automatic response to garden-pathing, contrary to what Frazier and Rayner originally assumed.

Garden-path sentences are rare in natural language, yet the SRH assumes that the parser automatically stores and uses spatial and linguistic information to guide regressive eye movements, which may not be realistic. Instead, selective reanalysis may need to be grouped with more specialized, controlled processes, as it is comparatively effortful, strategic, and would need to respond flexibly to properties of the stimulus at hand (Bugg & Crump, 2012). Cognitive control is likely to be correlated with awareness, though the two should not be equated (Hommel, 2007; Kunde, Reuss, & Kiesel, 2012). Based on this observation, we propose a new hypothesis that links selective reanalysis to conscious awareness of garden-pathing, which we term the *consciousness hypothesis* (CH):

Consciousness Hypothesis (CH): Reanalysis is more likely to be selective when it is conscious as opposed to automatic.

Suggestive evidence for the consciousness hypothesis comes from the P600 component in ERP research, which is linked to garden-pathing and other types of syntactic integration difficulty (e.g., Kaan, Harris, Gibson, & Holcomb, 2000), as well as to error detection in language comprehension more generally (Gouvea, Phillips, Kazanina, & Poeppel, 2010; van de Meerendonk, Kolk, Chwilla, & Vissers, 2009). The P600 has also been linked to awareness of syntactic violations (Batterink & Neville, 2013), as well as to regressive eye movements in response to such violations (Metzner, von der Malsburg, Vasishth, & Rösler, 2017), but a connection to *selective* regressions and conscious garden-pathing has, to our knowledge, not been established. Nevertheless, the link assumed by the consciousness hypothesis would be broadly consistent with this line of research.

Godfroid, Winke, and Rebuschat (2015) discuss the role of conscious awareness and control in selective reanalysis. They conclude that existing studies on native speakers are uninformative in this regard because they did not include explicit measures of awareness, such as verbal reports. Given the lack of prior evidence, it is helpful to first lay out how conscious reanalysis would unfold, and why it would be selective. Marcus (1980) describes conscious reanalysis as follows: When the parser is unable to parse an ambiguous sentence, the reader becomes aware of having been garden-pathed. A conscious “higher level problem solver” activates a set of heuristics and attempts to assign a syntactic structure without a full reparse (e.g., “If there is a verb without an accompanying subject, attempt to identify a misparsed relative clause in the preceding

input”). Marcus’s description implies selective attention to specific parts of the sentence, which is consistent with the consciousness hypothesis.² Different heuristics may be used by different speakers for different ambiguities, and new heuristics may conceivably be learned ad-hoc over the course of an experiment. Crucially, if conscious reanalysis is selective, it should be possible to infer the heuristics that a given reader is using from their reading patterns.

One way to test the consciousness hypothesis would be post-trial or post-study verbal reports, a method that has found much use in research on children and second-language learners (e.g., Eilers, Tiffin-Richards, & Schroeder, 2018, see Godfroid, Winke, & Rebuschat, 2015; Hessel, Nation, & Murphy, 2020 for reviews). While it is highly informative to ask subjects about their processing strategies (Afflerbach & Johnston, 1984), readers may not always have insight into the more subtle aspects of their own reading behavior (Hessel et al., 2020). Given this limitation, we use what we deem to be a more direct test of the relationship between consciousness and selectivity. Our paradigm, an extension of the well-known masked self-paced reading method (Just, Carpenter, & Woolley, 1982), allows participants to “regress” manually by pressing a specific key on the computer keyboard. Given that unspedded, uncued key presses are voluntary motor actions (e.g., Buehner, 2015; Hughes, Schütz-Bosbach, & Waszak, 2011), any rereading in such a paradigm can assumed to be conscious, and thus any indication of selective reanalysis would indicate conscious reanalysis. Besides answering the main research question (“Is there evidence for selective reanalysis when regressions are consciously controlled?”), another, more basic aim of the study was to investigate whether the proposed method is suitable for investigating sentence processing phenomena, and whether it can tell us anything that eye tracking and self-paced reading cannot.

2. The method: Bidirectional self-paced reading

In self-paced reading (SPR), sentences are divided into chunks (words or larger units), and the reader presses a key to display the chunks one by one. The time between the key presses is recorded. In the masked, non-cumulative “moving window” version, all letters of both upcoming and previous chunks are replaced with dashes (---). This manipulation is intended to prevent participants from revealing several chunks in succession and only then beginning to read them (Just et al., 1982), which prevents the localization of processing effects.

Regressions are not possible in standard self-paced reading. Given that non-negligible amounts of regressions do occur in normal reading, the method is therefore less naturalistic than eye tracking (Jegerski, 2014). One straightforward modification is to allow subjects to

² In their discussion of the diagnosis model of reanalysis, Fodor and Inoue (1994) also offer a detailed description of the “deductions” performed by the parser during reanalysis, but add that subjects are not assumed to be consciously aware of them.

return to earlier chunks. In our version, they can regress through the sentence in the same way they progress, namely by pressing a keyboard key to move from chunk to chunk. The resulting “bidirectional” self-paced reading (BSPR) paradigm has, to our knowledge, only been previously used by Kemtes, Brennan, and Wingfield (2001) and by Hatfield (2016), who implemented a touch-controlled version. Given the straightforwardness of the method, we speculate that it has otherwise been largely ignored because of the unnatural way regressions are performed. In effect, the sentence has to be read “backwards” to arrive at the target chunk. This is a strategy that Frazier and Rayner (1982) considered possible as a response to garden-pathing during free reading (“backward reanalysis”), but did not observe in their data.

It could be argued that forcing subjects into “backward reading” reduces the ecological validity of bidirectional SPR to a point where the method should not be used. However, subjects are not forced to reread every chunk immediately as they encounter it: they can “skip” chunks by simply pressing the relevant key again until they arrive at their regression target, assuming there is one. In addition, programming a target for a regressive episode should be easier than during free reading. Note that a crucial implicit assumption of the selective reanalysis hypothesis is that readers maintain a record of where previously read words are located in the display. Previous research on regressive eye movements indicates that readers can do this under specific experimental conditions, allowing them to selectively refixate material (Kennedy, 1982; Kennedy, Brooks, Flynn, & Prophet, 2003). However, the record appears to decay relatively rapidly and eye movements are not always precise (Inhoff & Weger, 2005), which may further contribute to the elusiveness of selective reanalysis effects. Bidirectional SPR eliminates both oculomotor noise and the need for spatial coding of word positions. Additionally, contaminating “time-out” regressions (Engelmann, Vasishth, Engbert, & Kliegl, 2013; Mitchell et al., 2008) are unlikely to occur, as it is implausible that participants would stall their progression by pressing a key to return to the previous region.

We now present the results of two studies on garden-path processing using web-based bidirectional self-paced reading. While BSPR allows much less freedom in participants’ trajectory through the sentence compared to free reading, and there can only be one “fixation” per visit to a region, it is still possible to compute many of the familiar reading measures from eye tracking: regions may be visited once or multiple times, regressions may be launched during the first pass or later, and even “skips” may occur when regions are passed through very rapidly. For our present purposes, we select two reading measures that can also be computed for eye-tracking data, namely right-bounded reading time and regressive rereading time, which together should give an impression of immediate, incremental processing and rereading after having viewed downstream material. In addition, we also conduct scanpath analyses. Scanpaths offer a more holistic way of looking at the eye movements within a given trial, and can be more informative

than single-region measures when investigating selective reanalysis (Christianson et al., in prep.; von der Malsburg & Vasishth, 2011, 2013). In order to maximize our chances of observing selective reanalysis, we rewarded both reading speed and comprehension accuracy with monetary payouts. The rationale behind this scheme was that adopting a selective rereading strategy would simultaneously fulfill the speed goal by minimizing unnecessary rereading and the accuracy goal by aiding syntactic reanalysis and thus correct interpretation.

Our studies did not aim to directly compare trials in which participants engage in conscious versus automatic reanalysis, due to the aforementioned difficulty of reliably assessing conscious awareness on a trial-by-trial basis. We also did not aim to directly compare BSPR to eye tracking or standard SPR, given the differences between the paradigms, though comparisons with effect sizes from earlier work are still possible. Rather, we aimed to create a setting that should, assuming that the consciousness hypothesis holds, be maximally conducive to selective reanalysis. Viewed against the mixed nature of previous results from natural reading, which is (presumably) largely automatic, observing robust selective rereading patterns in such a setting would be tentative evidence in favor of the consciousness hypothesis. By contrast, failure to observe selective reanalysis would add to the overall empirical picture suggesting that selective reanalysis is rare or possibly does not exist as a general mechanism (Christianson et al., in prep.).

3. Experiment 1

3.1 Participants and materials

100 German native speakers below the age of 50 (29 female, 71 male; mean age 30) were recruited on Prolific (<https://www.prolific.co>; Palan & Schitter, 2018). They were paid £5 for their participation. Additional prescreening was applied such that participants were required to have successfully completed 80% of previous experiments they had signed up for on Prolific.

Two sets of garden-path sentences each were constructed. For each sentence, both an early disambiguation version and a late disambiguation (garden-path) version were created. The first set contained the well-known NP coordination/S coordination ambiguity that exists in both German and English. Speakers initially prefer the NP-coordination analysis in these structures, and are garden-pathed when the structure is disambiguated towards sentential coordination (e.g., Frazier, 1987). An example sentence is shown in (2). Diamonds (◊) indicate chunk boundaries. In the early disambiguation condition, the NP *the rebel* is masculine and carries unambiguous nominative case marking, which identifies it as the subject of a new clause. By contrast, in the late disambiguation condition, the NP is feminine and its case marking is ambiguous between nominative and accusative, thus allowing for an initial NP coordination misanalysis. Reanalysis should occur at *guarded*, which needs a subject.

(2) **Coordination ambiguity**

Die **Opposition** ◊ **versteckte** ◊ **den** **gesuchten Sympathisanten** ◊ **und** ◊
 the opposition hid the wanted sympathizer and
 [der **Rebell** / **die** **Rebellin**] (NP) ◊ [mit der **Machete**]
 the.NOM rebel.MASC the.NOM|ACC rebel.FEM with the machete
 (mod) ◊ [**bewachte**] (verb) ◊ **den Eingang** ◊ **des** **geheimen Lagers** ◊ ...
 guarded the entrance of.the secret camp
 ‘The opposition hid the wanted sympathizer and the rebel with the machete guarded
 the entrance of the secret camp ...’

The second set of sentences contained an ambiguity that does not exist in English, namely a subject-object ambiguity based on syncretism between nominative and accusative case for feminine noun phrases. German native speakers preferably analyze clause-initial case-ambiguous noun phrases as nominative-marked subjects and experience a garden path when the clause is disambiguated towards object-initial word order through number agreement on the verb (e.g., Hemforth, 1993; Meng & Bader, 2000a; Paape, Hemforth, & Vasishth, 2018). An example sentence is shown in (3). In the early disambiguation condition, the NP *the animal keeper* is masculine and overtly marked for accusative case, indicating an OVS structure. By contrast, in the late disambiguation condition, the NP is feminine and its case marking is ambiguous between nominative and accusative, allowing for an initial SVO misanalysis. Reanalysis should occur at *sniffed (at) and licked*, as the verbs are marked for plural and the would-be subject is singular.

(3) **Subject-object ambiguity**

Den Besuchern ◊ **war es schon** **aufgefallen:** ◊ [**Den** **Tierpfleger**
 the visitors was it already noticed the.ACC animal.keeper.MASC
 / **Die** **Tierpflegerin**] (NP) ◊ [**der/die** **das wohlriechende**
 / the.NOM|ACC animal.keeper.FEM who.MASC/FEM the fragrant
Kastanienshampoo benutzte,] (mod) ◊ [**beschnupperten und**
 chestnut.shampoo used sniffed.at.PL and
beschleckten] (verb) ◊ **die Tiere** ◊ **im Zoo** ◊ **besonders gern** ◊ ...
 licked.PL the animals in.the zoo especially gladly
 ‘The visitors had already noticed: The animal keeper who used the fragrant chestnut
 shampoo was sniffed at and licked by the animals in the zoo especially gladly ...’

For both sentence types, selective reanalysis as proposed by Frazier and Rayner (1982) predicts regressions from the verb region to the NP region. This prediction arises because the initial incorrect attachment of the noun phrases in the late-disambiguation (garden-path) conditions constitutes the error that reanalysis should aim to selectively correct. We have no specific

prediction as to which of the two constructions should be more difficult to reanalyze. The main aim of the study is not to compare the two constructions, but to test the generality of selective reanalysis. Corpus data collected by Hoeks, Hendriks, Vonk, Brown, and Hagoort (2006) for the NP coordination versus S coordination in Dutch, which is typologically very close to German, and by Bader and Koukouloti (2018) for German SVO versus OVS word order indicate quite similar frequency ratios of about 4.5/1 in favor of NP coordination and SVO word order, respectively. Nevertheless, the two structures may differ in their diagnosis and revision costs: Bader (2000) has argued that non-canonical OVS structures such as (3) are especially difficult to reanalyze compared to SVO structures, as required for the S coordination analysis in (2). Furthermore, Meng and Bader (2000a) have argued that the unexpected number agreement on the verb in (3) is not a particularly effective cue for reanalysis, because the morphological mismatch does not directly signal what the correct structure should be. Given these observations, it is possible that subject-object sentences create a stronger garden-path effect than coordination sentences.³

During the experiment, each subject read a total of 32 garden-path sentences drawn from a set of 64 via a Latin-squares procedure. Due to a classification error, there were 31 coordination ambiguity sentences and 33 subject-object ambiguity sentences. Fillers consisted of 20 sentences with varied structures, in addition to a set of 64 “memory” sentences, which all contained long-distance anaphoric references, e.g., “Johann had been given three five-Euro bills by his dad and a twenty-Euro bill by his mom to go shopping in the city. Unfortunately, he lost the money that his dad had given him on the bus”. Each subject read 32 of these memory sentences, which were designed to encourage rereading, resulting in a total of 84 sentences.

The garden-path stimuli were designed “with variability in mind” (Yarkoni, 2019): In order to increase the generalizability of the findings to a larger variety of natural language contexts, varying types of modifiers were used (relative clauses, prepositional phrases) and the absolute position of the critical regions as well as the distances between regions were varied between items. Additionally, some items consisted of multiple sentences while others consisted of a single complex sentence. Sentence content was varied to maximize engagement, and each stimulus was designed as a mini-discourse telling a coherent story. The full list of garden-path stimuli with their accompanying comprehension questions is given in the online appendix at <https://osf.io/j8cbh>.

3.2 Procedure and data analysis

The experiment was run on the Ixet farm Drummond (2013). Participants were instructed to read the sentences silently and answer the accompanying comprehension questions. They were told to press the right arrow key to move forward through the sentence and to press the left

³ Note, however, that the disambiguating region in subject-object sentences is also longer than in coordination sentences, which may lead to more pronounced effects.

arrow key to move backward through the sentence. They were also informed that they could press the ESC key at any point to return directly to the beginning of the sentence, and the CTRL key to continue directly to the comprehension question.

A comprehension question was asked after every trial. Comprehension questions were detailed and targeted various aspects of the sentence, including the critical dependency in garden-path sentences in a small minority of cases, e.g., “How much money did Johann lose?”, “Where did the visitors see the animal keeper?”, “Did the opposition hide someone?”. Half of all comprehension questions were yes/no questions while the other half were multiple-choice questions with three possible answers each. Participants were instructed not to draw inferences beyond the literal content of the sentences when answering the questions.

Participants were told that there would be an anonymous high-score list, and that the ten highest scorers out of 100 would be paid an additional bonus of £9 each. They were told that they would gain 2 points for each correct answer to a comprehension question and that they would automatically lose 1 point every two minutes until the end of the experiment. After giving informed consent, participants completed two unscored practice trials, after which they were reminded of the scoring procedure and the function of each keyboard key. Afterwards, the main experiment began.

Due to their parallel structure, both kinds of garden-path sentences were analyzed together. Trials in which any single-region reading time was longer than 5000 ms were removed, resulting in a loss of 2% of the data. We chose right-bounded reading time (RBRT), regressive rereading time (RRT), and the proportion of first-pass regressions from a region as our measures of interest. Right-bounded reading time includes all reading times on the region prior to moving on to the next region; it is thus composed of first-pass reading time and progressive rereading time, that is, time spent rereading the region after having made a regression but before moving on to the right. By contrast, regressive rereading time includes all reading times on the region *after* having moved on to the right; it includes regressive as well as progressive rereading occurring after this point. Garden-pathing in the late disambiguation conditions compared to the early disambiguation conditions should lead to increased right-bounded reading times and possibly more first-pass regressions at the disambiguating verb region. Under the selective reanalysis hypothesis, regressive rereading times at the ambiguous NP region should be increased in the late disambiguation conditions.

All regions with RBRT and/or RRT below 150 ms were removed from the respective analyses, under the rationale that these very short RTs are the BSPR equivalent of skipping a region. This procedure resulted in the removal of 1% of the data for RBRT, and of about 80% of the data for RRT, mainly due to trials with RRT = 0 ms (that is, absence of any regressive rereading) being removed. Note that this means that the RRT analysis quantitatively compares the time spent rereading in the 20% of trials in which rereading above 150 ms actually occurred within the

window of analysis. It does not capture differences between conditions in whether any rereading occurred at all. This aspect of the data is covered by the scanpath analysis presented below.

We defined a window of analysis ranging from the early disambiguation region – that is, the noun phrase at the beginning of the ambiguous region in the garden-path conditions – to two regions after the disambiguating verb.⁴ Linear mixed-effects models (LMMs) with lognormal likelihoods were fitted to individual regions using the `brms` package for Bayesian inference (Bürkner, 2017, 2018) in R (R Core Team, 2020). The models included the sum-coded predictors garden-path type (–1 for coordination, 1 for subject-object) and disambiguation (–1 for early, 1 for late). The full analysis code, including prior specifications, is given in the online appendix.

3.3 Reading measures: Descriptive results and LMM analysis

Figure 1 shows log-transformed RBRT and RRT by region. Subject-object sentences were read more slowly than coordination sentences across regions, conditions, and measures, likely due to the different clause structures, but we focus on the effects of disambiguation and its interactions. We report percentile-based 95% credible intervals of effects for which 95% of the posterior probability are above or below zero. Note that this is merely a convention we adopt for reporting effects for which there is some support in the data, and does not correspond to a null-hypothesis significance test. When interpreting the results, the reader should consider the width of the credible intervals, whether a given interval includes zero, and whether the effect replicates (see below). The full range of results is given in the online appendix.

3.3.1 Right-bounded reading times

Early disambiguation at the NP region affected reading in coordination and subject-object sentences differently (interaction: $\hat{\Delta} = -38$ ms, CrI: [–61 ms, –15 ms]). Coordination sentences showed an ambiguity-induced slowdown in the NP region, that is, longer reading times in the late disambiguation condition ($\hat{\Delta} = 31$ ms, CrI: [2 ms, 62 ms]) while subject-object sentences showed a slowdown due to disambiguation, that is, faster reading times in the late disambiguation condition ($\hat{\Delta} = -44$ ms, CrI: [–78 ms, –10 ms]). At the verb region, there was an indication of a garden-path effect ($\hat{\Delta} = 23$ ms, CrI: [2 ms, 45 ms]), and no indication of an interaction with garden-path type. The garden-path effect continued into the verb + 1 region ($\hat{\Delta} = 15$ ms, CrI: [–4 ms, 34 ms]) and the verb + 2 region ($\hat{\Delta} = 13$ ms, CrI: [–2 ms, 30 ms]).

⁴ In subject-object sentences, an adverb sometimes intervened between the verb region and the subject noun phrase. In these cases, the subject noun phrase was coded as the verb + 1 region. For items in which the noun modifier consisted of multiple regions, these were collapsed and the mean reading time was used. One item did not contain a modifier, so it was dropped from the relevant analyses.

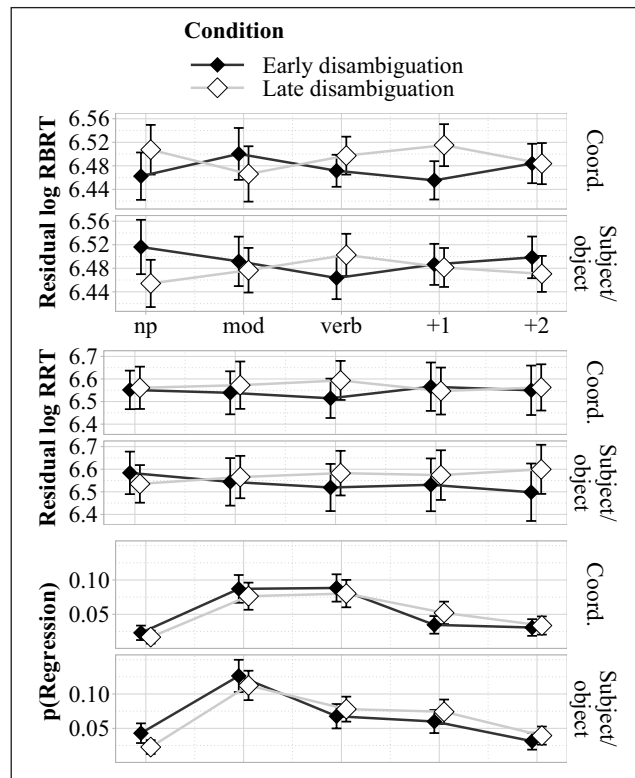


Figure 1: Reading measures for Experiment 1. Due to different region lengths, log reading measures have been residualized against garden-path type and region ID to improve visibility. RBRT = Right-bounded reading time. RRT = Regressive rereading time. p (Regression) = Proportion of first-pass regressions.

3.3.2 Regressive rereading times

There was no indication of garden-path effects or interactions with garden-path type.

3.3.3 First-pass regressions

There was no indication of garden-path effects or interactions with garden-path type.

3.4 Scanpath analysis

Scanpaths were analyzed using the Scasim algorithm developed by von der Malsburg and Vasishth (2011), which was designed for eye-tracking data. Scasim computes scanpath dissimilarities based on both fixation locations and fixation durations: If scanpaths have fixations in different locations and/or fixations of different durations in the same location, they are more dissimilar than scanpaths with fixations matching in both location and duration. The scanpath analysis complements the analysis of reading measures in multiple ways: First, it is holistic, in the sense that the entire trial is always considered instead of a single region. Second, it can capture

differences in rereading probabilities that were not considered in the region-based analysis. Third, by dividing the scanpath similarities by the total duration of the trial, we obtain a measure of how attention was distributed across different regions of the sentence (see below).

In order to imitate eye-tracking data as closely as possible, we created x - and y -coordinates for each region of interest by setting y to a constant value n and x to n multiplied by the region number, thus simulating horizontal movement by increments of n . Here, we set n to 50. Region-wise reading times were used as fixation durations. For the scanpath analysis, reading times below 150 ms (“skips”) were retained, as they are an integral part of the scanpath’s overall shape. The reading times were log-transformed and residualized against region length, age of the participant, and the absolute position of the region within the sentence, including random effects by subject and by item. All effects had t -values >2 . The residualization procedure is necessary because Scasim is blind to hierarchical structure in the data: Large differences between “fixations” can be due to extraneous factors, such as individual differences in reading speed and position-related speedups (e.g., Demberg & Keller, 2008; Ferreira & Henderson, 1993), lowering the chance of identifying the systematic reading patterns we are interested in.

In order to make scanpaths comparable across sentences, only the following regions were considered: the beginning of the sentence, the NP (early disambiguation) region, the immediately following modifier region(s), the verb (late disambiguation) region, the following verb + 1 region, and the end of the sentence.⁵ Any “fixations” prior to the first visit to the NP region were discarded. The left panel of **Figure 2** shows the critical portion of the scanpath predicted by selective reanalysis: When the disambiguating verb is encountered, a series of regressions should be triggered that targets the NP, quickly crossing over the modifier. After reanalysis has occurred, forward reading should resume, presumably without additional processing of subsequent regions. The right panel of **Figure 2** shows another possible scanpath that would also show some indication of selectivity: The participant reads the entire sentence, then quickly returns to the beginning of the sentence, continues forward, carries out reanalysis at the NP, and then again proceeds forward. This pattern could be called “selective forward reanalysis”. By contrast, a more unselective rereading pattern would show approximately equal amounts of time spent in each region. Due to its sensitivity to differences in region-wise reading times, the scanpath analysis should be able to distinguish the two patterns.

Scasim was set to normalize to total scanpath durations to obtain similarity per unit of time, as suggested by von der Malsburg and Vasisht (2011). The resulting similarity measure can be interpreted as the difference in the amount of attention paid to each region as a proportion of the

⁵ We did not consider the verb + 2 region in the scanpath analysis, as it contained no theoretically interesting material. Furthermore, the proportion of regressions from this region was very low, and including it in the analysis tended to decrease the fit of the scanpath model while yielding overall similar scanpath patterns.

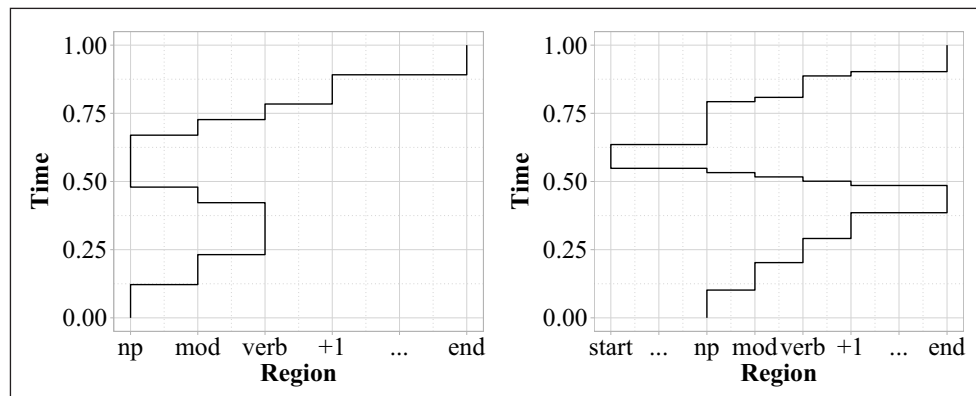


Figure 2: Left: Partial scanpath predicted by selective reanalysis. Right: Hypothetical “selective forward reanalysis” scanpath. Note the increased rereading times at the NP region in both scanpaths. The y-axis shows the time spent in each region as a proportion of the total duration of the trial.

duration of the entire trial. The matrix of pairwise scanpath similarities produced by Scasim was transformed into a two-dimensional Euclidean map using nonmetric multidimensional scaling as implemented in the *vegan* package (Oksanen et al., 2020). Given that multidimensional scaling algorithms are nondeterministic and may converge to different local stress optima,⁶ we repeated the procedure several times and chose the configuration with the lowest stress and the best interpretability (Borg & Mair, 2017). Mixture-based clustering was applied to the result using the *mclust* package (Scrucca, Fop, Murphy, & Raftery, 2016), and the number of clusters was reduced using the entropy criterion of Baudry, Raftery, Celeux, Lo, and Gottardo (2010).

The two dimensions of the resulting scanpath space can be roughly interpreted as “amount of rereading” and “location of regressions”. The scanpath map had a stress value of 11%. As shown in **Figure 3**, five clusters emerged: A large cluster of scanpaths with extensive rereading but no selective pattern (cluster 1, “extensive rereading”), two clusters of scanpaths with no rereading at all (clusters 2 and 3, “no rereading”), a cluster with single-region regressions into the NP region (cluster 4, “mod-to-NP”), and finally a cluster with single-region regressions from the verb region (cluster 5, “verb-to-mod”). **Figure 4** shows example scanpaths for each of the clusters.

For each cluster, we fitted LMMs with Bernoulli likelihoods to the membership of the observed scanpaths (in cluster versus not in cluster). The models included condition (early disambiguation versus late disambiguation), garden-path type (coordination versus subject-object) and participants’ centered speed/accuracy point scores as fixed effects, as well as all two-way interactions with condition. Clusters 2 and 3 (“no rereading”) were collapsed for this analysis, as we could discern no meaningful difference between the patterns. Compared to early

⁶ Stress is a measure of how well the transformation preserves the original similarities, with 0% stress being the theoretical absolute optimum.

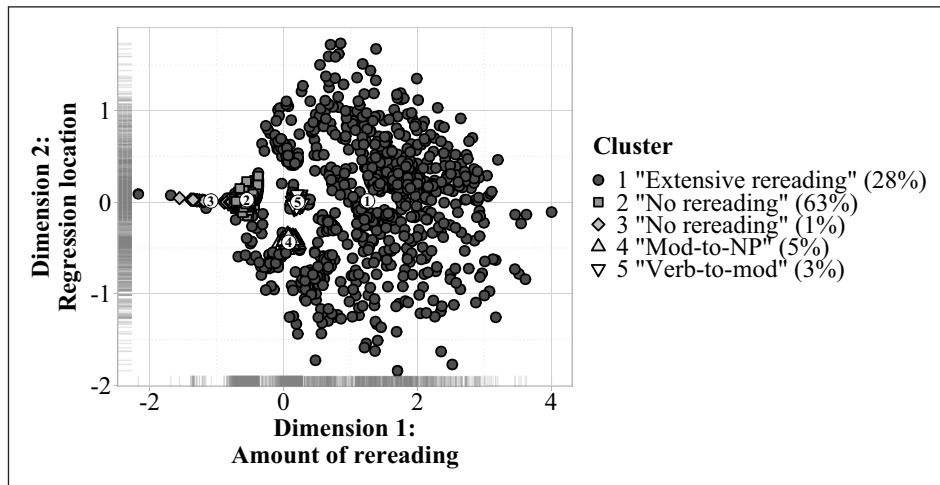


Figure 3: Scanpath map from Experiment 1. Note that the scanpaths in all clusters except cluster 1 (“extensive rereading”) are very uniform, which leads to overplotting. The rug plots in the margins show the densities along the two dimensions of the scanpath space.

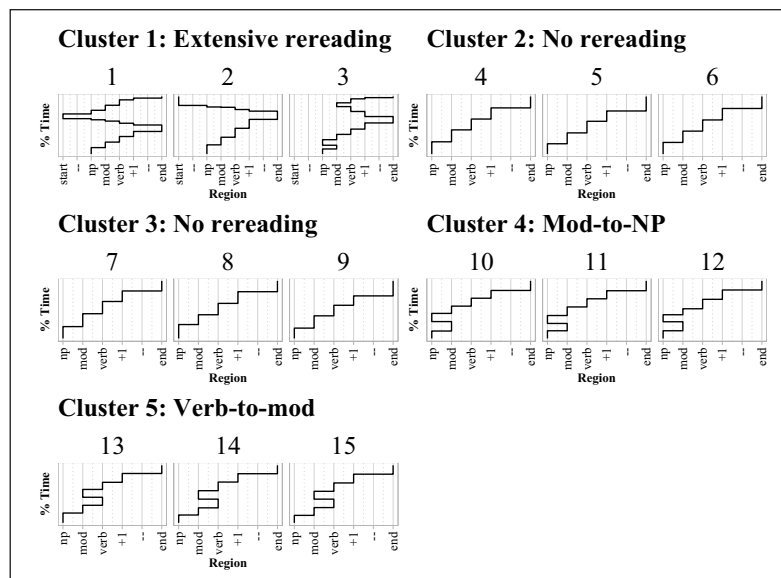


Figure 4: Three example scanpaths for each of the five clusters from Experiment 1. The regions are as follows: 1 – Beginning of sentence, 2 — (removed), 3 – NP (early disambiguation), 4 – Post-NP, 5 – Verb (late disambiguation), 6 – Post-verb, 7 – (removed), 8 – End of sentence.

disambiguation, late disambiguation increased the probability of scanpaths falling into cluster 1, the “extensive rereading” cluster ($\hat{\Lambda} = 4\%$, CrI: [1%, 8%]), and decreased the probability of scanpaths falling into clusters 2 and 3, the “no rereading” clusters ($\hat{\Lambda} = -5\%$, CrI: [-9%, -1%]). For cluster 1 (“extensive rereading”), there was also an interaction between disambiguation and score ($\hat{\Lambda} = 3\%$, CrI: [-1%, 6%]), such that high-scoring readers were especially likely to show

extensive rereading in the late-disambiguation conditions ($\hat{\Delta} = 10\%$, CrI: $[-2\%, 22\%]$) compared to the early-disambiguation conditions ($\hat{\Delta} = 3\%$, CrI: $[-7\%, 14\%]$).

Compared to coordination sentences, subject-object sentences had more scanpaths falling into cluster 4 (“mod-to-NP”), irrespective of disambiguation ($\hat{\Delta} = 2\%$, CrI: $[0\%, 3\%]$). Participants with higher scores were somewhat less likely to show this regression pattern ($\hat{\Delta} = -1\%$, CrI: $[-2\%, 0\%]$). There were no effects in evidence regarding membership in cluster 5 (“verb-to-mod”).

We investigated the relationship between cluster membership and question response accuracy by fitting LMMs to question response accuracies using cluster as a predictor. The predictor was treatment-coded, with the two “no rereading” clusters serving as the baseline. Across all clusters, participants’ mean question response accuracy was 83%. Across participants, accuracies ranged from 41% to 100% (1st quartile: 75%, 3rd quartile: 94%).⁷ Results of the analysis showed that scanpath membership in cluster 1 (“extensive rereading”) was associated with higher question response accuracy compared to clusters 2 and 3 (“no rereading”), irrespective of disambiguation ($\hat{\Delta} = 4\%$, CrI: $[1\%, 6\%]$).

Lastly, in order to investigate how the varied rereading patterns within the “extensive rereading” cluster were affected by the experimental manipulations, we fitted LMMs with Gaussian likelihoods to both dimensions of the scanpath space of cluster 1. The models contained the categorical predictors disambiguation and garden-path type, as well as two continuous predictors: the total number of points scored by the participant in the experiment and the position of the trial in the experiment. Both continuous predictors were centered and scaled. The models also contained all two-way interactions with disambiguation.

Compared to early disambiguation, late disambiguation increased rereading on average ($\hat{\Delta} = 0.05$, CrI: $[-0.01, 0.11]$). Participants with higher scores also engaged in more rereading ($\hat{\Delta} = 0.17$, CrI: $[0.06, 0.28]$) and showed more varying regression locations ($\hat{\Delta} = 0.05$, CrI: $[-0.01, 0.11]$).

3.5 Discussion

Our results suggest that bidirectional self-paced reading can detect garden-path effects during right-bounded reading, that is, first-pass reading plus progressive rereading. Using scanpath analyses, recurring patterns can be identified in participant’s rereading strategies – or non-strategies.

Rereading was, by assumption, consciously controlled in this experiment. Under these conditions, we found no evidence of selective rereading, which should have manifested as a cluster of regressions from the verb region to the NP region in the garden-path conditions.

⁷ Two participants performed below chance on the garden-path sentences, but overall showed above-chance performance with regard to question response accuracy.

Extensive, seemingly unselective rereading, however, occurred more often in the garden-path condition, and was tied to higher question response accuracy when compared against single-pass reading, across both garden-path and control sentences. There were also scanpath clusters with single, short regressions to an immediately preceding region. Such clusters were also observed by von der Malsburg and Vasishth (2011, 2013), who speculated that these regressions were tied to “checking” of word forms and/or corrections of premature forward saccades. Both explanations could also apply to our data. However, as the clusters were small and did not respond to the garden-path manipulation, we refrain from speculating about the precise function of these regressions and the possible role of awareness in their generation.

Our method also detected individual differences between participants. Participants with higher scores on the speed-accuracy measure across the entire experiment engaged in more rereading of late-disambiguated garden-path sentences, despite the fact that time spent rereading was penalized. It appears that these participants were able to generate a net benefit in comprehension accuracy through rereading, though there was no indication that this benefit was limited to garden-path sentences. We return to this point later.

We discuss the broader significance of the findings, especially with regard to selective reanalysis and the consciousness hypothesis, in section 5. We postpone the discussion because we were interested in whether our findings are replicable before judging their theoretical implications. We therefore conducted a direct replication study with new participants, using the same method and the same sentences, and ran the same statistical analyses on the new sample.

4. Experiment 2

4.1 Participants and materials

A new sample of 100 German native speakers below the age of 50 was recruited (36 female, 64 male; mean age 30). Payment and prescreening were the same as in Experiment 1.

4.2 Procedure and data analysis

Procedure and data analysis were identical to Experiment 1. Again, trials in which any single-region reading time was longer than 5000 ms were removed, resulting in a loss of 1.6% of the data. Trials with reading measures below 150 ms (including RRT = 0 ms) were also removed, resulting in a loss of 1% the data for RBRT and 83% of the data for RRT.

4.3 Reading measures: Descriptive results and LMM analysis

Figure 5 shows log-transformed RBRT and RRT by region. As before, we focus on the effects of disambiguation and its interactions. The full range of results is given in the online appendix.

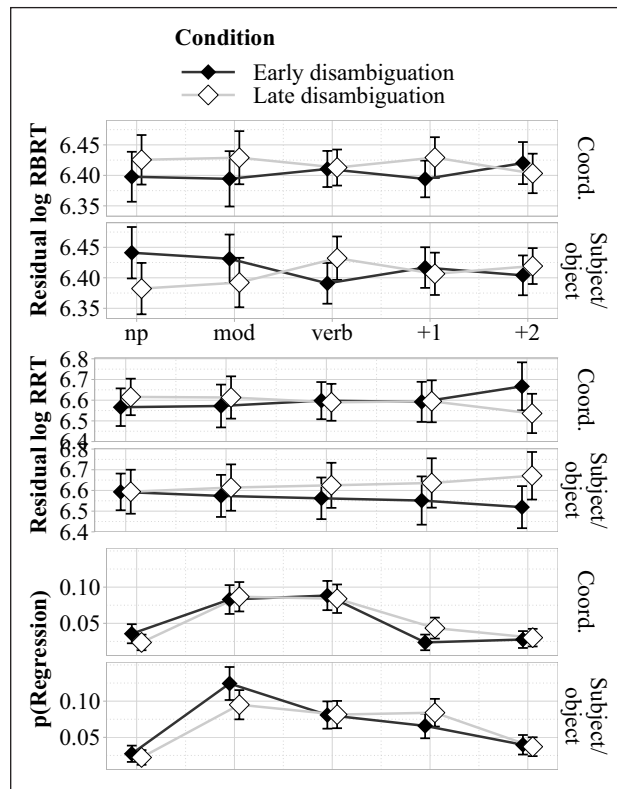


Figure 5: Reading measures for Experiment 2. Due to different region lengths, log reading measures have been residualized against garden-path type and region ID to improve visibility. RBRT = Right-bounded reading time. RRT = Regressive rereading time. p (Regression) = Proportion of first-pass regressions.

4.3.1 Right-bounded reading times

As in Experiment 1, there was an interaction between disambiguation and garden-path type in the NP region (interaction: $\hat{\Delta} = -24$ ms, CrI: [-46 ms, -2 ms]), which was mainly driven by late-disambiguated subject-object sentences being read faster than early-disambiguated ones ($\hat{\Delta} = -34$ ms, CrI: [-68 ms, -1 ms]). At the verb region, there was an indication of a garden-path effect ($\hat{\Delta} = 17$ ms, CrI: [-5 ms, 38 ms]). As **Figure 5** shows, the effect was mainly driven by subject-object sentences, and correspondingly there was some indication of an interaction with garden-path type ($\hat{\Delta} = 16$ ms, CrI: [-4 ms, 36 ms]).

4.3.2 Regressive rereading times

At the verb+2 region, there was an indication of an interaction between disambiguation and garden-path type ($\hat{\Delta} = 56$ ms, CrI: [-3 ms, 115 ms]): in the late disambiguation compared to the early disambiguation condition, subject-object sentences tended to have longer regressive rereading times while coordination sentences tended to have shorter regressive rereading times.

Given that the verb + 2 region contained no material that was relevant to resolving the garden path, we speculate that the effect is probably spurious.

4.3.3 First-pass regressions

At the verb + 1 region, there were more first-pass regressions in the late disambiguation conditions compared to the early disambiguation conditions ($\hat{\Delta} = 1\%$, CrI: [0%, 3%]).

4.4 Scanpath analysis

As in Experiment 1, scanpaths clustered along the dimensions “amount of rereading” and “location of regressions”. The scanpath map had a stress value of 10%. As shown in **Figure 6**, five clusters emerged, most of which corresponded to the ones observed in Experiment 1: A cluster of scanpaths with extensive rereading (cluster 1, “extensive rereading”), a large cluster of scanpaths without any rereading (cluster 2, “no rereading”), a cluster with short regressions into the NP region (cluster 5, “mod-to-NP”), and a cluster with short regressions from the verb region to the preceding NP modifier region (cluster 4, “verb-to-mod”). A cluster that had not emerged in Experiment 1 showed short regressions into the verb region (cluster 3, “post-verb-to-verb”). **Figure 7** shows example scanpaths for each of the clusters.

As in Experiment 1, garden-pathing increased the probability of membership in cluster 1, the “extensive rereading” cluster, though the effect was less pronounced ($\hat{\Delta} = 1\%$, CrI: [-2%, 5%]). Unlike in Experiment 1, scanpaths of participants with higher scores were more likely to belong to this cluster ($\hat{\Delta} = 14\%$, CrI: [5%, 25%]), and there was also an interaction between

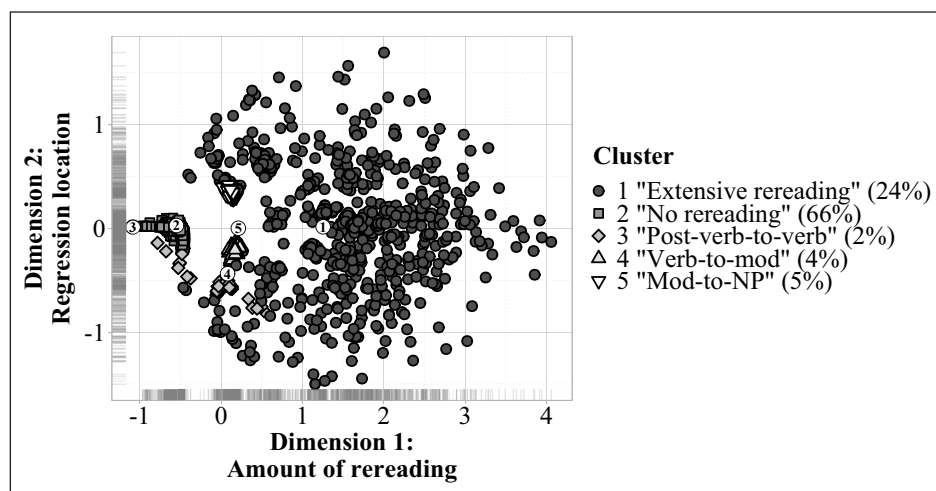


Figure 6: Scanpath map from Experiment 2. Note that the scanpaths in all clusters except cluster 1 (“extensive rereading”) are very uniform, which leads to overplotting. The rug plots in the margins show the densities along the two dimensions of the scanpath space.

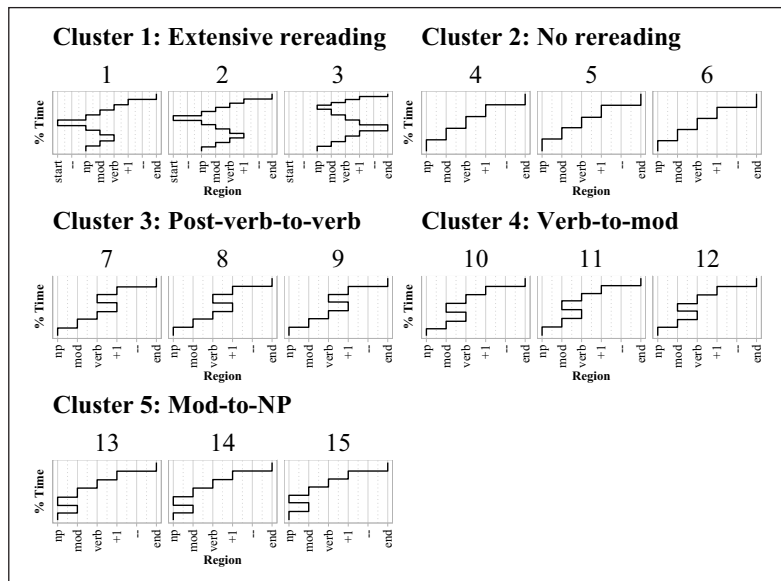


Figure 7: Three example scanpaths for each of the five clusters from Experiment 2. The regions are as follows: 1 – Beginning of sentence, 2 – (removed), 3 – NP (early disambiguation), 4 – Post-NP, 5 – Verb (late disambiguation), 6 – Post-verb, 7 – (removed), 8 – End of sentence.

score and disambiguation ($\hat{\Delta} = -4\%$, CrI: $[-8\%, -1\%]$): Cluster membership was more affected by score in the early-disambiguation conditions ($\hat{\Delta} = 19\%$, CrI: $[9\%, 31\%]$) compared to the late-disambiguation conditions ($\hat{\Delta} = 10\%$, CrI: $[0\%, 21\%]$). Participants with higher scores were less likely to show scanpaths belonging to cluster 2, the “no rereading” cluster ($\hat{\Delta} = -19\%$, CrI: $[-32\%, -6\%]$). As in Experiment 1, scanpaths in subject-object sentences were more likely to fall into cluster 5 (“mod-to-NP”), irrespective of disambiguation ($\hat{\Delta} = 2\%$, CrI: $[0\%, 3\%]$).

Participants’ mean question response accuracy was 82%. Across participants, accuracies ranged from 25% to 100% (1st quartile: 75%, 3rd quartile: 94%).⁸ As in Experiment 1, scanpath membership in cluster 1 (“extensive rereading”) was associated with higher question response accuracy compared to cluster 2 (“no rereading”), irrespective of disambiguation ($\hat{\Delta} = 5\%$, CrI: $[2\%, 7\%]$). Furthermore, membership in cluster 4 (“verb-to-mod”) was associated with higher question response accuracy ($\hat{\Delta} = 6\%$, CrI: $[0\%, 8\%]$). The latter effect is likely due to task demands, as comprehension questions often targeted information from the modifier. Scanpath membership in cluster 5 (“mod-to-NP”) increased accuracy in the early-disambiguation conditions ($\hat{\Delta} = 6\%$, CrI: $[0\%, 8\%]$) but not in the late-disambiguation conditions ($\hat{\Delta} = 0\%$, CrI: $[-13\%, 6\%]$). Given that early disambiguation occurred on the NP, this pattern could indicate successful “checking” of the NP’s disambiguating case feature.

⁸ Four participants performed below chance on the garden-path sentences, but overall showed above-chance question response accuracy.

As in Experiment 1, we analyzed the two scanpath dimensions within cluster 1 (“extensive rereading”). For the rereading dimension, late disambiguation sentences again showed more rereading compared to early disambiguation sentences ($\hat{\Lambda} = 0.06$, CrI: [-0.02, 0.15]). As in Experiment 1, participants with higher scores engaged in more extensive rereading ($\hat{\Lambda} = 0.09$, CrI: [0, 0.18]). There was also an interaction between disambiguation and trial number ($\hat{\Lambda} = 0.07$, CrI: [0, 0.14]), such that rereading decreased with trial number for early-disambiguation sentences ($\hat{\Lambda} = -0.11$, CrI: [-0.21, 0]) but not for late-disambiguation sentences ($\hat{\Lambda} = 0.03$, CrI: [-0.08, 0.13]). This interaction may be spurious, as the opposite tendency was observed in Experiment 1. Unlike in Experiment 1, regression location did not seem to be affected by participant score ($\hat{\Lambda} = -0.02$, CrI: [-0.06, 0.03]).

4.5 Discussion

The most relevant findings from Experiment 1 were replicated in Experiment 2. As in Experiment 1, the two garden-path types diverged at the early disambiguation region. For both garden-path types, there was a garden-path effect at the late disambiguation region in right-bounded reading times. Unlike in Experiment 1, the garden-path effect was also visible in first-pass regressions from the post-verbal region. Again, the scanpath analysis showed that garden-pathing led to more extensive rereading on average. There was again no indication of the rereading pattern predicted by selective reanalysis when rereading was manually controlled. However, extensive rereading was again tied to higher question response accuracy across conditions.

The early disambiguation penalty at the NP that we observed for unambiguous object-first sentences in both experiments is well-attested in German (Hemforth, 1993; Konieczny, 1996; Paape et al., 2018). The acceptability of OVS word order in German partly depends on contextual licensing conditions (Verhoeven & Temme, 2017; Weskott, Hörnig, Fanselow, & Kliegl, 2011) which may not always have been fully met in our items, despite prescreening for subjective acceptability by two native speakers. By contrast, the early ambiguity cost observed for coordination sentences may reflect competition between the NP coordination and S coordination analyses. However, earlier studies on the coordination ambiguity in Dutch did not show a competition effect (Hoeks et al., 2006; Hoeks, Vonk, & Schriefers, 2002). We suspect that the pattern in our studies may have been incurred by a clash between subjects’ underlying preference for a complex NP analysis and the chunking we imposed. As the potential complex NP was broken up (*The opposition* ◊ hid ◊ the wanted sympathizer ◊ and ◊ the rebel ... ◊ guarded ...), subjects may have been pushed towards an S-coordination analysis. Given that our main conclusions do not depend on this pattern, we will not speculate on it in more depth.

We now turn to the implications of our findings for the consciousness hypothesis, as well as the more general implications for bidirectional self-paced reading as an experimental paradigm.

5. General discussion

The aim of our experiments was twofold: First, we wanted to investigate the consciousness hypothesis, which states that reanalysis in garden-path sentences is more likely to be selective when it is consciously controlled. Second, and more generally, we wanted to investigate whether bidirectional self-paced reading (BSPR) is a paradigm that warrants further exploration.

Our results indicate that bidirectional self-paced reading can detect garden-path effects. Aggregating across both experiments ($n_{total} = 200$) and both sentence types (coordination and subject-object ambiguity), the estimated garden-path effects in right-bounded reading times at the late disambiguation/verb region is 19 ms (CrI: [4 ms, 34 ms]). Across experiments, at the post-disambiguation region, there was also a 1% increase in regressions (CrI: [0%, 2%]) in the late compared to the early disambiguation conditions. The effect on regressions is small, but the same numerical pattern was visible across experiments and across both garden-path types. **Figure 8** shows the estimates for right-bounded reading times and first-pass regressions across experiments.

Our estimates for right-bounded reading times are within the plausible range of effects for coordination ambiguities from eye tracking, self-paced reading, and self-guided reading, where participants reveal words by moving a slider to the left or right (Hatfield, 2016). In their eye-tracking study, Witzel, Witzel, and Forster (2012) observed a marginally significant difference of 23 ms⁹ in first-pass reading times and of 33 ms in right-bounded reading times. In their self-paced reading study, they observed a non-significant difference of 52 ms. Hatfield (2016) reports an estimate of 26 ms (CI: [1 ms, 52 ms]) in total reading times from self-guided reading, and an estimate of 6 ms (CI: [-19 ms, 32 ms]) for self-paced reading. Based on these comparisons, we conclude that BSPR is, in principle, a suitable method for investigating online ambiguity resolution.

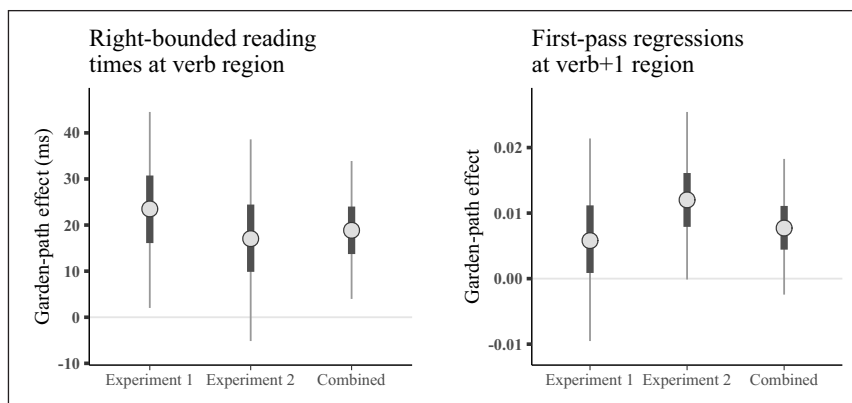


Figure 8: Comparison of estimates for the effects of late versus early disambiguation across experiments. Inner bars correspond to 50% credible intervals, outer bars correspond to 95% credible intervals.

⁹ As the authors do not report standard errors, no confidence interval can be calculated.

5.1 Implications for the consciousness hypothesis, and for selective reanalysis more generally

Our results do not lend credibility to the consciousness hypothesis, at least under the assumption that every key press in BSPR is the result of a conscious decision. We believe that this assumption is mostly warranted, especially for regressive key presses. For standard self-paced reading, it has been argued that subjects may start using a “tapping” strategy, advancing through the sentence mechanically at a fixed speed (Witzel et al., 2012). It is possible that the “no regression” clusters seen in our scanpath analyses correspond to such a strategy. Still, it is doubtful that readers would change direction without conscious awareness. When readers did decide to reread, regression patterns were varied and influenced by the experimental manipulations. Unselective rereading was more common in the late disambiguation conditions, that is, in the garden-path conditions. This matches earlier findings from eye tracking with Spanish sentences by von der Malsburg and Vasishth (2011, 2013), who also found more “unselective” rereading in the garden-path condition.

Our data do not indicate that readers focused on the ambiguous region of the sentence when they were garden-pathed: Across both experiments, there was no indication of increased regressive rereading times in the ambiguous NP region ($\hat{\Delta} = -9$ ms, CrI: [-55 ms, 36 ms]), nor was there a scanpath cluster that would have indicated targeted revisits of the ambiguous NP region starting from the disambiguating verb region. The results thus do not lend support to the prediction that conscious “double takes” upon disambiguation coincide with selective reanalysis. By contrast, our findings are compatible with the view that selective reanalysis is not a general mechanism employed by the human sentence processor (Christianson et al., in prep.). However, given the assumed conscious nature of rereading the BSPR paradigm, it can also be argued that our results are compatible with the assumptions made by Frazier and Rayner (1982): Frazier and Rayner argued that conscious reanalysis is *not* selective, because it indicates a failure of automatic parsing routines. The absence of evidence for selective reanalysis in our studies could thus be taken to indicate a mixture of different types of trials: reanalysis was either easy enough to be carried out covertly (Lewis, 1998; von der Malsburg & Vasishth, 2013), or it was difficult, to an extent where both covert reanalysis and selective reanalysis failed, resulting in conscious, unselective rereading.

We do not know how often readers succeeded at reanalyzing our garden-path structures, as most comprehension questions did not target the relevant dependency. It is also possible that readers resorted to “good enough” processing in some proportion of trials, sticking to their incorrect initial interpretations (e.g., Christianson, Hollingworth, Halliwell, & Ferreira, 2001), or simply decided that the sentences were malformed, without even trying to engage in reanalysis (Fodor & Inoue, 2000; Meng & Bader, 2000b). These issues can be investigated in future work by using only comprehension questions that target the critical dependencies, or by using a grammaticality judgment task.

In our data sets, efficient readers – that is, readers who were fast and showed accurate comprehension across stimulus types – were more likely to engage in extensive, “unselective” rereading, and this type of rereading was tied to higher question response accuracy. While the pattern was not exclusive to garden-path sentences, it suggests that extensive rereading is, perhaps unsurprisingly, a useful strategy for information extraction. This may be especially true for sentences such as ours, which, in an attempt to increase naturalness, provided somewhat more context and plausibility than typical stimuli used in psycholinguistic research.

5.2 Generalizability of the results

Given the large number of participants, we could presumably have obtained narrower credible intervals for our effects, had we not chosen to increase the variability in our stimulus materials to aid generalizability. In our experiments, we obtained estimates of the same magnitude for the standard deviations of the slope adjustments to the garden-path effects by sentence and by subject (that is, by-item and by-subject “random slopes”). In a typical psycholinguistic experiment, the observed variability between sentences can be an order of magnitude smaller than the variability between subjects, because sentence structure is tightly controlled and real-life linguistic variation is eliminated (Yarkoni, 2019). This reduces noise and increases statistical power, but also reduces naturalness and generalizability. Our stimuli were, of course, still not natural in the sense of having been sampled from real-life language and occurring within a larger discourse (Hamilton & Huth, 2020), but presumably somewhat closer to participants’ everyday experience.

One important question related to generalizability is whether our findings extend to free reading. Despite its doubtful relationship with comprehension accuracy (Christianson et al., in prep.), “unselective” rereading may be the preferred strategy for handling garden paths in free reading (von der Malsburg & Vasishth, 2011, 2013), which matches our findings. Regressions in free reading are not fully automatic, as indicated by the fact that fewer regressions are made when previously read text is made unavailable for rereading through masking (Booth & Weger, 2013; Christianson et al., in prep.; Schotter et al., 2014; White, Lantz, & Paterson, 2017), presumably because there is little to gain from a regression. Furthermore, regression behavior during unmasked reading is influenced by task demands such as time pressure and the difficulty of comprehension questions (Godfroid, Loewen, et al., 2015; Weiss, Kretzschmar, Schlesewsky, Bornkessel-Schlesewsky, & Staub, 2018; Wotschack & Kliegl, 2013). As stated in the introduction, our task demands were tailored to favor selective rereading. Across our two experiments, subjects made at least one regression in garden-path sentences (early and late disambiguation) in about 47% of trials. The median number of regressions for these trials was 3, up to a maximum of 42. Participants jumped forward to the comprehension question in about 6% of trials and returned directly to the beginning of the sentence in about 5% of trials. The resulting variability in scanpath patterns is visualized in **Figure 9**, which shows a sample of scanpaths taken from cluster 1 (“extensive rereading”) of Experiment 2.

Note that despite our use of the term “unselective” for this cluster of patterns, there are clear subpatterns showing rereading of the entire sentence (trial 2061), as well as trials showing the expected regression pattern of selective reanalysis (trial 2060), though without much indication of any prolonged rereading of the ambiguous NP region. In future work, we plan to examine whether comparable results can be obtained in an eye-tracking paradigm with the same task demands. Another open question that we plan to address is whether our findings from bidirectional self-paced reading generalize to other types of garden paths across languages. For instance, based on eye-tracking results, neither selective nor “unselective” rereading appears to increase comprehension accuracy for English NP/Z sentences (*Since Jay always jogs a mile and a half seems ...*; Christianson et al., 2017, in prep.); this pattern may change when rereading is consciously controlled.

5.3 The need to consider conscious reading strategies

As a more general conclusion, we believe that research on syntactic ambiguity resolution and other psycholinguistic phenomena would benefit from closer engagement with the literature on metacognitive reading strategies that supplement automatic, lower-level reading skills (Afflerbach, Pearson, & Paris, 2008). Research in this domain often focuses on educational aspects, such as teaching efficient reading strategies to poor readers (Mokhtari & Reichard, 2002). Furthermore, the proposed strategies, such as taking notes or underlining parts of the text, usually target text-level comprehension, and are difficult to transfer to the sentence level and the modalities of a typical psycholinguistic experiment. Other strategies, however, may be highly relevant, such as the adjustment of reading speed according to text difficulty, use of contextual

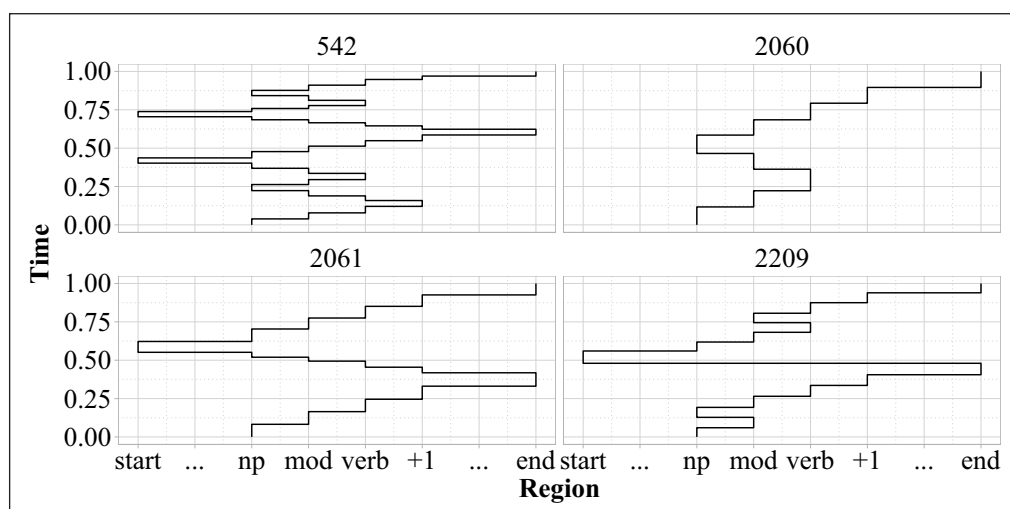


Figure 9: Example scanpaths for cluster 1 (“extensive rereading”) from Experiment 2. The regions are as follows: 1 – Beginning of sentence, 2 — (removed), 3 – NP (early disambiguation), 4 – Post-NP, 5 – Verb (late disambiguation), 6 – Post-verb, 7 – (removed), 8 – End of sentence.

clues, active comprehension monitoring, and, most importantly in the present context, rereading (Mokhtari & Reichard, 2002). At the text level, strategic rereading improves comprehension, metacomprehension, and information recall (Millis & King, 2001; Rawson, Dunlosky, & Thiede, 2000; Stine-Morrow, Gagne, Morrow, & DeWall, 2004), but texts containing known triggers of processing difficulty such as negation may be exempt from this advantage (Margolin & Snyder, 2018). Given the results of Christianson et al., garden-path effects may be another example of this pattern.

Text-level reading strategies engage cognitive control mechanisms (Moss, Schunn, Schneider, McNamara, & VanLehn, 2011), which are also implicated in garden-path resolution (Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2014; Novick, Trueswell, & Thompson-Schill, 2005; Woodard, Pozzan, & Trueswell, 2016), as well as in the triggering of regressions during reading (Luke & Henderson, 2013). In this context, one interesting question for future research is whether encountering a disambiguating word in a garden-path sentence should be considered an explicit control trigger or an implicit control trigger. An example of an explicit control trigger would be a pre-trained visual or auditory cue that directly tells the participant to stop performing the current task (Logan, 1982). By contrast, implicit control triggers must be learned from the environment, such as the frequency of incongruent Stroop trials in a given experiment (Kunde et al., 2012). It would appear that disambiguating information in garden-path sentences is an implicit control trigger, given that the participant is not explicitly told to reanalyze the syntactic structure. The distinction is important with regard to selectivity, because implicit triggers of cognitive control have been argued to require awareness to be effective (Kunde et al., 2012). This would imply that controlled reanalysis – selective or unselective – cannot be carried unless the garden path is consciously registered. More generally, it remains to be seen if garden-path recovery is best characterized as a case of goal pursuit in the absence of conscious awareness (Custers & Aarts, 2010), or whether awareness has some role to play after all.

6. Conclusion

We have shown that despite its limitations compared to free reading, bidirectional self-paced reading is a promising method for sentence processing research, especially for the study of phenomena that may involve conscious awareness. Assuming that self-paced reading is always consciously controlled, we found no evidence that conscious rereading in garden-path sentences is selective. While this result is in line with the claim of Frazier and Rayner (1982) that selective reanalysis does not involve awareness, our study adds to a growing body of research suggesting that by default, rereading as a response to garden-pathing is unselective rather than selective.

Abbreviations

FEM — Feminine

MASC — Masculine

NOM — Nominative

ACC — Accusative

PL — Plural

NP — Noun phrase

Supplementary files

The experimental stimuli, the data, and our analysis code are available at <https://osf.io/j8cbh>.

Ethics and consent

The experiments were conducted in accordance with the Declaration of Helsinki, as last revised. Informed consent was obtained from all participants prior to experimentation, based on a detailed description of the experimental procedure, the reward scheme, and our use of the submitted data.

Acknowledgements

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

Funding information

The experiments were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 428960187 – PI: Dario Paape.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

DP acquired funding for the project, supervised the project and was responsible for its administration. DP and SV conceptualized the experiment and decided on the methodology. DP took the lead on data collection, data analysis, data curation, and writing. SV reviewed, edited, and authorized the text. Program code was written by DP and SV.

References

- Afflerbach, P., & Johnston, P. (1984). On the use of verbal reports in reading research. *Journal of Reading Behavior*, 16 (4), 307–322. DOI: <https://doi.org/10.1080/10862968409547524>
- Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, 61 (5), 364–373. DOI: <https://doi.org/10.1598/RT.61.5.1>
- Bader, M. (2000). On reanalysis: Evidence from German. In B. Hemforth & L. Konieczny (Eds.), *German sentence processing* (pp. 187–246). Dordrecht: Springer Netherlands. DOI: https://doi.org/10.1007/978-94-015-9618-3_7
- Bader, M., & Koukouloti, V. (2018). When object-subject order is preferred to subject-object order: The case of German main and relative clauses. In E. Fuß, M. Konopka, B. Trawinski, & U. H. Waßne (Eds.), *Grammar and corpora* (pp. 53–71). Heidelberg: Heidelberg University Publishing.
- Batterink, L., & Neville, H. J. (2013). The human brain processes syntax in the absence of conscious awareness. *Journal of Neuroscience*, 33 (19), 8528–8533. DOI: <https://doi.org/10.1523/JNEUROSCI.0618-13.2013>
- Baudry, J.-P., Raftery, A. E., Celeux, G., Lo, K., & Gottardo, R. (2010). Combining mixture components for clustering. *Journal of Computational and Graphical Statistics*, 19 (2), 332–353. DOI: <https://doi.org/10.1198/jcgs.2010.08111>
- Booth, R. W., & Weger, U. W. (2013). The function of regressions in reading: Backward eye movements allow rereading. *Memory & Cognition*, 41 (1), 82–97. DOI: <https://doi.org/10.3758/s13421-012-0244-y>
- Borg, I., & Mair, P. (2017). The choice of initial configurations in multidimensional scaling: Local minima, fit, and interpretability. *Austrian Journal of Statistics*, 46 (2), 19–32. DOI: <https://doi.org/10.17713/ajs.v46i2.561>
- Buehner, M. J. (2015). Awareness of voluntary and involuntary causal actions and their outcomes. *Psychology of Consciousness: Theory, Research, and Practice*, 2 (3), 237–252. DOI: <https://doi.org/10.1037/cns0000068>
- Bugg, J. M., & Crump, M. J. (2012). In support of a distinction between voluntary and stimulus-driven control: A review of the literature on proportion congruent effects. *Frontiers in Psychology*, 3 , 367. DOI: <https://doi.org/10.3389/fpsyg.2012.00367>
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80 (1), 1–28. DOI: <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10 (1), 395–411. DOI: <https://doi.org/10.32614/RJ-2018-017>
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42 (4), 368–407. DOI: <https://doi.org/10.1006/cogp.2001.0752>
- 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 (7), 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)
- Custers, R., & Aarts, H. (2010). The unconscious will: How the pursuit of goals operates outside of conscious awareness. *Science*, 329 (5987), 47–50. DOI: <https://doi.org/10.1126/science.1188595>
- Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109 (2), 193–210. DOI: <https://doi.org/10.1016/j.cognition.2008.07.008>
- Devitt, M. (2006). Intuitions in linguistics. *The British Journal for the Philosophy of Science*, 57 (3), 481–513. DOI: <https://doi.org/10.1093/bjps/axl017>
- Drummond, A. (2013). Ibex 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 modelling the interaction of syntactic processing and eye movement control. *Topics in Cognitive Science*, 5 (3), 452–474. DOI: <https://doi.org/10.1111/tops.12026>
- Ferreira, F., & Henderson, J. M. (1993). Reading processes during syntactic analysis and reanalysis. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 47 (2), 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 (5), 407–434. DOI: <https://doi.org/10.1007/BF02143947>
- Fodor, J. D., & Inoue, A. (2000). Garden path repair: Diagnosis and triage. *Language and Speech*, 43 (3), 261–271. DOI: <https://doi.org/10.1177/00238309000430030201>
- Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language & Linguistic Theory*, 5 (4), 519–559. DOI: <https://doi.org/10.1007/BF00138988>
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14 (2), 178–210. DOI: [https://doi.org/10.1016/0010-0285\(82\)90008-1](https://doi.org/10.1016/0010-0285(82)90008-1)
- Godfroid, A., Loewen, S., Jung, S., Park, J.-H., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition*, 37 (2), 269–297. DOI: <https://doi.org/10.1017/S0272263114000850>
- Godfroid, A., Winke, P., & Rebuschat, P. (2015). 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>
- Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, 25 (2), 149–188. DOI: <https://doi.org/10.1080/01690960902965951>

- Hamilton, L. S., & Huth, A. G. (2020). The revolution will not be controlled: Natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, 35 (5), 573–582. DOI: <https://doi.org/10.1080/23273798.2018.1499946>
- Hatfield, H. (2016). Self-guided reading: Touch-based measures of syntactic processing. *Journal of Psycholinguistic Research*, 45 (1), 121–141. DOI: <https://doi.org/10.1007/s10936-014-9334-2>
- Hemforth, B. (1993). *Kognitives Parsing: Repräsentation und Verarbeitung sprachlichen Wissens*. Sankt Augustin: Infix.
- Hessel, A. K., Nation, K., & Murphy, V. A. (2020). Comprehension monitoring during reading: An eye-tracking study with children learning English as an additional language. *Scientific Studies of Reading*, 25 (2), 159–178. DOI: <https://doi.org/10.1080/10888438.2020.1740227>
- Hoeks, J. C., Hendriks, P., Vonk, W., Brown, C. M., & Hagoort, P. (2006). Processing the noun phrase versus sentence coordination ambiguity: Thematic information does not completely eliminate processing difficulty. *Quarterly Journal of Experimental Psychology*, 59 (9), 1581–1599. DOI: <https://doi.org/10.1080/17470210500268982>
- Hoeks, J. C., Vonk, W., & Schriefers, H. (2002). Processing coordinated structures in context: The effect of topic-structure on ambiguity resolution. *Journal of Memory and Language*, 46 (1), 99–119. DOI: <https://doi.org/10.1006/jmla.2001.2800>
- Hommel, B. (2007). Consciousness and control: Not identical twins. *Journal of Consciousness Studies*, 14 (1–2), 155–176.
- Hughes, G., Schütz-Bosbach, S., & Waszak, F. (2011). One action system or two? evidence for common central preparatory mechanisms in voluntary and stimulus-driven actions. *Journal of Neuroscience*, 31 (46), 16692–16699. DOI: <https://doi.org/10.1523/JNEUROSCI.2256-11.2011>
- Inhoff, A. W., & Weger, U. W. (2005). Memory for word location during reading: Eye movements to previously read words are spatially selective but not precise. *Memory & Cognition*, 33 (3), 447–461. DOI: <https://doi.org/10.3758/BF03193062>
- Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology*, 69 (5), 907–925. DOI: <https://doi.org/10.1080/17470218.2014.984231>
- Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 20–49). New York: Routledge. DOI: <https://doi.org/10.4324/9780203123430>
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111 (2), 228–238. DOI: <https://doi.org/10.1037/0096-3445.111.2.228>
- Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15 (2), 159–201. DOI: <https://doi.org/10.1080/016909600386084>
- 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>

- Kennedy, A. (1982). Eye movements and spatial coding in reading. *Psychological Research*, 44 (4), 313–322. DOI: <https://doi.org/10.1007/BF00309327>
- Kennedy, A., Brooks, R., Flynn, L.-A., & Prophet, C. (2003). The reader's spatial code. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 193–212). Amsterdam: Elsevier. DOI: <https://doi.org/10.1016/B978-044451020-4/50012-8>
- Konieczny, L. (1996). *Human sentence processing: A semantics-oriented parsing approach* (Unpublished doctoral dissertation). University of Freiburg.
- Kunde, W., Reuss, H., & Kiesel, A. (2012). Consciousness and cognitive control. *Advances in Cognitive Psychology*, 8 (1), 9–18. DOI: <https://doi.org/10.5709/acp-0097-x>
- 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. (1982). On the ability to inhibit complex movements: A stop-signal study of typewriting. *Journal of Experimental Psychology: Human Perception and Performance*, 8 (6), 778–792. DOI: <https://doi.org/10.1037/0096-1523.8.6.778>
- Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 13 (2), 123–146. DOI: <https://doi.org/10.1080/1057356970130203>
- Luke, S. G., & Henderson, J. M. (2013). Oculomotor and cognitive control of eye movements in reading: Evidence from mindless reading. *Attention, Perception, & Psychophysics*, 75 (6), 1230–1242. DOI: <https://doi.org/10.3758/s13414-013-0482-5>
- Marcus, M. P. (1980). *A theory of syntactic recognition for natural language*. Cambridge, MA: MIT Press.
- Margolin, S. J., & Snyder, N. (2018). It may not be that difficult the second time around: the effects of rereading on the comprehension and metacomprehension of negated text. *Journal of Research in Reading*, 41 (2), 392–402. DOI: <https://doi.org/10.1111/1467-9817.12114>
- Meng, M., & Bader, M. (2000a). Mode of disambiguation and garden-path strength: An investigation of subject-object ambiguities in German. *Language and Speech*, 43 (1), 43–74. DOI: <https://doi.org/10.1177/00238309000430010201>
- Meng, M., & Bader, M. (2000b). Ungrammaticality detection and garden path strength: Evidence for serial parsing. *Language and Cognitive Processes*, 15 (6), 615–666. DOI: <https://doi.org/10.1080/016909600750040580>
- Meseguer, E., Carreiras, M., & Clifton, C. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, 30 (4), 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>

- Millis, K. K., & King, A. (2001). Rereading strategically: The influences of comprehension ability and a prior reading on the memory for expository text. *Reading Psychology*, 22 (1), 41–65. DOI: <https://doi.org/10.1080/02702710151130217>
- 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 (3), 266–293. DOI: <https://doi.org/10.1016/j.jml.2008.06.002>
- Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, 94 (2), 249–259. DOI: <https://doi.org/10.1037/0022-0663.94.2.249>
- Moss, J., Schunn, C. D., Schneider, W., McNamara, D. S., & VanLehn, K. (2011). The neural correlates of strategic reading comprehension: Cognitive control and discourse comprehension. *NeuroImage*, 58 (2), 675–686. DOI: <https://doi.org/10.1016/j.neuroimage.2011.06.034>
- 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 (2), 186–217. DOI: <https://doi.org/10.1080/01690965.2012.758297>
- Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 5 (3), 263–281. DOI: <https://doi.org/10.3758/CABN.5.3.263>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2020). *vegan: Community ecology package* [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=vegan> (R package version 2.5-7)
- Paape, D., Hemforth, B., & Vasishth, S. (2018). Processing of ellipsis with garden-path antecedents in french and german: Evidence from eye tracking. *PLOS ONE*, 13 (6), 1–46. DOI: <https://doi.org/10.1371/journal.pone.0198620>
- 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>
- 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 (4), 940–961. DOI: <https://doi.org/10.1037/0278-7393.24.4.940>
- Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago, IL: University of Chicago Press.
- R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Rawson, K. A., Dunlosky, J., & Thiede, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition*, 28 (6), 1004–1010. DOI: <https://doi.org/10.3758/BF03209348>

- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105* (1), 125–157. DOI: <https://doi.org/10.1037/0033-295X.105.1.125>
- 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* (6), 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* (1), 289–317. DOI: <https://doi.org/10.32614/RJ-2016-021>
- Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingerin misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69* (2), 104–120. DOI: <https://doi.org/10.1016/j.jml.2013.04.001>
- Stine-Morrow, E. A., Gagne, D. D., Morrow, D. G., & DeWall, B. H. (2004). Age differences in rereading. *Memory & Cognition*, *32* (5), 696–710. DOI: <https://doi.org/10.3758/BF03195860>
- Sturt, P. (1996). Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes*, *11* (5), 449–494. DOI: <https://doi.org/10.1080/016909696387123>
- Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*, *40* (1), 136–150. DOI: <https://doi.org/10.1006/jmla.1998.2606>
- van de Meerendonk, N., Kolk, H. H., Chwilla, D. J., & Vissers, C. T. W. (2009). Monitoring in language perception. *Language and Linguistics Compass*, *3* (5), 1211–1224. DOI: <https://doi.org/10.1111/j.1749-818X.2009.00163.x>
- Verhoeven, E., & Temme, A. (2017). Word order acceptability and word order choice. In S. Featherston, R. Hörnig, R. Steinberg, B. Umbreit, & J. Wallis (Eds.), *Proceedings of linguistic evidence 2016 – empirical, theoretical, and computational perspectives*. Tübingen: Universität Tübingen.
- von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, *65* (2), 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* (10), 1545–1578. DOI: <https://doi.org/10.1080/01690965.2012.728232>
- Weiss, A. F., Kretzschmar, F., Schlesewsky, M., Bornkessel-Schlesewsky, I., & Staub, A. (2018). Comprehension demands modulate re-reading, but not first-pass reading behavior. *Quarterly Journal of Experimental Psychology*, *71* (1), 198–210. DOI: <https://doi.org/10.1080/17470218.2017.1307862>
- Weskott, T., Hörnig, R., Fanselow, G., & Kliegl, R. (2011). Contextual licensing of marked OVS word order in German. *Linguistische Berichte*, *2011* (225), 3–18.

- White, S. J., Lantz, L. M., & Paterson, K. B. (2017). Spontaneous rereading within sentences: Eye movement control and visual sampling. *Journal of Experimental Psychology: Human Perception and Performance*, 43 (2), 395–413. DOI: <https://doi.org/10.1037/xhp0000307>
- Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: Eye tracking, moving-window, and maze. *Journal of Psycholinguistic Research*, 41 (2), 105–128. DOI: <https://doi.org/10.1007/s10936-011-9179-x>
- Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology*, 141, 187–209. DOI: <https://doi.org/10.1016/j.jecp.2015.08.005>
- Wotschack, C., & Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *Quarterly Journal of Experimental Psychology*, 66 (3), 548–562. DOI: <https://doi.org/10.1080/17470218.2011.625094>
- Yarkoni, T. (2019). The generalizability crisis. *PsyArXiv preprint*. DOI: <https://doi.org/10.31234/osf.io/jqw35>

