The role of working memory and punctuation in the processing of Garden Path sentences: An Online Study

Scheifer, Gilles R.

The role of working memory and punctuation in the processing of Garden Path sentences

An Online Study

Gilles R. Scheifer
M.A. Thesis Linguistics, Leiden University

Supervisor: Dr. L. Pablos Robles

Second Reader: Dr. S. Valdez

June 8th, 2022

Table of Contents

Acknowledgments

I want to thank the following people for helping me not only with my thesis but also across my studies:
My parents, **John Scheifer** and **Maggy Mersch**, for their ceaseless support over the years be it with respect to university or elsewhere.
Michel Max for helping me out in numerous scripting assignments and testing the experiment included in the current study.
Marc Weber and **Ben Karier** for testing the experiment included in the current study.
Laurent Langehegermann for listening to me talk about this thesis and my study in general over the years.
Dr. Leticia Pablos Robles for being my supervisor during this and the previous thesis and providing me with support.
Dr. Susana Valdez for taking the time to be the second reader for this thesis.

Abstract

Ambiguity resolution has been a topic of debate in language processing models. The present study investigated the impact of punctuation and working memory in Garden Path structures and related these concepts with assumptions made under the Good Enough parsing approach. A word-by-word self-paced reading task was used to assess the effects. A backwards digit span and a reading span task were administered to measure the participant's working memory capacity. The study found evidence for the presence of Garden Path effects in Late Closure structures. In the form of response accuracy to the comprehension questions, the data supports the Good Enough parsing approach's assumption that ambiguities are not fully resolved. The scores on the backwards digit span task correlated positively with response accuracy on comprehension questions while reading span scores did not. Punctuation did not impact reading time of the disambiguating region nor response accuracy.

1. Background

Sentence processing is an ubiquitous occurrence, however the processes underlying the rapid phenomenon are not yet fully understood. Over the past decades, a number of theories and models have been forwarded in an attempt to encapsulate all aspects of sentence processing, however none has garnered enough support to be generally accepted. Points of contention range from the material necessary to start building a sentence structure to the number of structures constructed in parallel. Initially, researchers hypothesised that new material was being integrated into the currently established sentence structure, resulting in a bottom-up approach . Over time, cognitive sciences altogether moved towards predictive approaches and compounded with counter-evidence mostly stemming from ERP components and eye-tracking studies, theories of sentence processing followed the trend (Ferreira & Chantavarin, 2018).

Within the predictive models of sentence processing, one major distinction which can be made between different approaches is whether the approach adheres to a serial- or parallel processing account. In order to discern different integration strategies, ambiguous sentences and their interpretations have been a focal point of experiments investigating sentence processing. Parallel models such as the constraint based model typically assumed that the difficulty in sentence processing stemmed from competition between multiple possible alternatives which were considered in parallel (McRae et al., 1998; van Gompel et al., 2000).

On the other hand, serial approach models posit that only one predictive structure is being built at a time, and that processing difficulties arise when the parser realises that the structure which is currently being built cannot accommodate the latest material. One of the most prominent serial processing models is the Garden Path approach (Frazier & Fodor, 1978)**.** While the initial model is dated and has been revised by the authors themselves (Frazier & Clifton, 1996, 1997), the concept of reanalysis has been provided additional support with the advent of the Good Enough parsing approach (Blache, 2016; Christianson et al., 2001, 2006; Ferreira et al., 2001; Ferreira & Patson, 2007; Traxler, 2014). Good Enough parsing postulates that the parser does not fully analyse the incoming material, but rather sacrifices precision to gain speed. While this trade-off is said to usually be efficient, in the case of temporarily ambiguous sentences, it has been found that the comprehension questions following them were frequently answered as if the first analysis was still in the participant's mind.

1.1 The Garden Path Approach

The Garden Path model is a serial sentence processing model, first proposed by Frazier & Fodor (1978). The Garden Path model is a two-stage model which proposes that in the first stage only syntactic information is being processed, which is subsequently enriched by lexical information during the second stage (Frazier & Fodor, 1978; Scheifer, 2021). It postulates that new information is gradually integrated into the currently established structure once encountered following three principles, which were derived from those established in Kimball (1973) (Frazier & Clifton, 1997).

(1) Late Closure

If grammatically permissible, attach new items into the close or phrase currently being processed.

(2) Minimal attachment

Do not postulate any potentially unnecessary nodes.

(3) Minimal Chain Principle

Postulate required chain members at the earliest point grammatically possible, but postulate no potentially unnecessary chain members (Vincenzi, 1991).

Together, the principles predict the parser to integrate new information into the current phrase if permitted by the grammar in such a way that no unnecessary nodes are being generated while keeping the chains at a minimal length (Frazier & Clifton, 1997; Frazier & Rayner, 1987; Kimball, 1973; Scheifer, 2021; Vincenzi, 1991). Evidence for this approach comes, among others, from Garden Path sentences. A Garden Path sentence is a sentence which is temporarily ambiguous until clearly disambiguating material is encountered later in the sentence. Gráf (2014) identified 10 types of Garden Path constructions, and likely more exist. One type of Garden Path sentence typically presents the parser with a second noun phrase which could plausibly serve as either the object of the preceding verb, only to later reveal that it has to be interpreted as the subject of the matrix clause which a parser only realises when being confronted with the second verb. If the reader were to follow the initial analysis, the second verb would end up without an agent role assigned and the resulting construction consequently ungrammatical. The example in (1) is a typical Garden Path sentence. (Gráf, 2014)

(4) *Example of a Garden Path sentence*

When Mary knitted the socks fell to the floor.

Late Closure expects the parser to interpret *the socks* as the object of the verb *knitted*, as it would be a grammatically acceptable position. However, upon encountering the second verb *fell*, the parser is forced to abandon the initial construction. This process of reevaluating the initial interpretation is known as reanalysis. Reanalysis has been observed both in reading time experiments (Christianson et al., 2006; Scheifer, 2021) and eye-tracking experiments. However, as the number of Garden Path studies increased, methodological concerns regarding the presentation of the Garden Path stimuli have arisen. One of these concerns regards the punctuation of the stimuli. As you may have noticed when reading the example in (1), the comma which is demanded by English grammar rules to separate the matrix- and sub clause has been omitted. This is typically done in studies investigating Garden Path effects to further mislead the reader as it promotes the interpretation that the second NP, *the socks* in (4), is the direct object of the preceding verb. Therewith, the reader can only realise the misguided interpretation when encountering the second verb, *fell* in (1), as it cannot be assigned a subject under the initially presumed sentence structure.

1.2 Previous study

Scheifer (2021) investigated different accounts of predictive processing using an online experiment to measure reading time data of native speakers of English encountering different degrees of ambiguities and filler-gap constructions.

The study found evidence supporting the Garden Path model (Frazier & Fodor, 1978) through the collected reading time data showing a significant increase in the temporarily ambiguous Garden

Path condition when compared to the fully disambiguous or disambiguating conditions.

The study also investigated the Hyperactive Gap Filling account (Omaki et al., 2015), however, was unable to find satisfactory evidence for the account, an element which would benefit from further research. This was potentially due to the paradigm being unable to check for the filler-gap constructions as hoped.

The Good Enough parsing approach was also investigated and the study found evidence for it in the form of a steep decline in response accuracy following comprehension questions relating to temporarily ambiguous sentences. Under Good Enough parsing, that result is indicative of the ambiguity not being fully resolved and the parser therefore failing to provide the correct response.

1.3 Punctuation and Garden Pathing

The benefit of omitting the comma is that the paradigms more consistently elicit the Garden Path effect as it has been previously found that punctuation can serve as a disambiguating variable in ambiguous sentences. This assumption was first borne out of the *Implicit Prosody Hypothesis* (Bader, 1998; Fodor, 1998, 2002) which postulates that during reading, the reader generates prosodic boundaries, which are helpful in interpreting the upcoming sentence elements. Evidence for this disambiguation has been convincing and one-sided for auditory stimuli, where prosodic breaks are used to disambiguate sentences (Hirotani et al., 2006; Kjelgaard & Speer, 1999; White et al., 2014). However, the effects of punctuation and its similarity to auditory prosodic boundaries as claimed by the *Implicit Prosody Hypothesis,* have been more disputed. Steinhauer & Friederici (2001), investigating German, found disambiguating effects through the insertion of commas which are hypothesised to serve as the written equivalent of a prosodic boundary. The study also investigated event related potentials, or ERP in short, which are electrophysiological brain responses measurable using an electroencephalogram. ERP components are of relevance as different kinds of stimuli elicit different responses from the brain, and those responses can be measured and differentiated by analysing their magnitude, direction and timing. Being able to quantify the brain's reaction to different kinds of stimuli allows researchers to draw conclusions about the nature of these stimuli. As such, if the insertion of a comma in a written stimulus elicits the same ERP component as a prosodic boundary in an auditory stimulus, it can be hypothesised that they are underlyingly treated as the same by the brain. For auditory prosodic boundaries, studies, such as Steinhauer (1999) on German, had found an ERP component which had become known as the Closure Positive Shift. This component can be characterised as a positive response following a prosodic boundary (Steinhauer, 2014). Steinhauer & Friederici (2001) elicited the CPS for visually presented stimuli which included commas, supporting the assumption that commas are the written equivalent of prosodic boundaries, and therewith supporting the *Implicit Prosody Hypothesis.* A study in English carried out by Drury et al. (2016) also found supporting evidence for the parallels between prosodic breaks and punctuations in visual stimuli in the form of the CPS. The study investigated the impact of commas on gender mismatch effects and ambiguities resulting from *Late Closure* (1). While commas seemed to influence parsing in both paradigms, the results surrounding *Late Closure* are of central interest for the present study. These findings suggest that the presence of a grammatically expected comma facilitates the correct interpretation of Late Closure type Garden Path sentences, while grammatically superfluous, additional commas reduced the acceptability of each sentence. On the other hand, in a study on Dutch Kerkhofs et al. (2008) failed to find the CPS ERP component when introducing commas into visually presented temporarily ambiguous sentences, while the identical stimuli did elicit the CPS when presented auditorily. This conflicting evidence points towards one or more yet unknown variables playing a role in disambiguating sentences.

1.4 The role of working memory

Individual differences are another consideration with rising prominence when it comes to sentence processing. The field of linguistics has started moving further and further away from the ideal speaker, and has taken an increased interest in the role of which individual differences play when it comes to linguistic performance, including sentence processing. One of the individual differences under discussion is working memory, how it works and which aspects of language processing are influenced by it. However, the findings regarding the influence of working memory on ambiguity resolution have been mixed. Two principal approaches to the role of working memory in language processing have sprung up. The *Dedicated Resource* account proposes a separation between interpretive and non-interpretive tasks. Interpretive tasks are those which construct the meaning of incoming linguistic information, and the *Dedicated Resource* account that these have some working memory capacity specifically allocated to them, while the non-interpretive tasks share the general working memory pool with other non-linguistic tasks. On the other hand, the *Shared Resources* accounts see sentence processing as sharing a limited working memory pool with all other tasks, be they linguistic or not (Brothers et al., 2021). *Shared Resources* accounts attribute a large role in sentence processing to the general working memory (Brothers et al., 2021; Fedorenko et al., 2006) , while the *Dedicated Resources* accounts claim other individual differences as more crucial for performance in sentence processing. (Brothers et al., 2021; Freed et al., 2017; Traxler, 2012; Traxler et al., 2005; Van Dyke et al., 2014)

Supporting the *Shared Resources* accounts, Kim & Christianson (2013) investigated the impact of working memory in temporarily ambiguous sentences for Korean and English. Their study found a significant effect of working memory in the disambiguating region for English with low working memory participants taking significantly longer to read the disambiguating material than the high working memory participants. In Korean, no such influence of working memory could be found, a difference which has been attributed to the different head-directionalities of the respective languages. Gordon et al. (2002) and subsequently, Fedorenko et al. (2006) used a dual-task paradigm where participants had to read and comprehend sentences while remembering different nouns. They manipulated the word type of the nouns to be remembered to either be the same or different from those used in the sentences read. Both studies found that participants performed worse on sentence comprehension when the word type was the same across the word list and the sentences. These findings were taken as support for the *Shared Resource* account, as it implies that sentence processing relies on working memory resources outside of the syntactic processing system (Gordon et al., 2002). Kim & Christianson (2017) have found that performance in self-paced reading tasks involving temporary ambiguities correlates negatively with working memory scores. This result has been attributed to high working memory parsers maintaining the different options presented by the ambiguity with more prominence leading to competition. This hypothesis is borne out of the findings of the study, with high working memory readers taking longer in the region where the ambiguous nature of the sentence became apparent, and the same readers taking longer to decide which of the options presented by the ambiguity to select. Cheng et al. (2021) investigated the impact of working memory on attachment resolution in L1 and L2 speakers to check whether the same underlying processes were being utilised. While working memory did have an impact on the offline task, with high performance individuals having a preference for low attachment, no significant effect was found on the online task.

A link between working memory capacity and Garden Path constructions comes from evidence by a study by Farmer et al. (2017). In this study, participants underwent a battery of working memory tasks, as well as a word-by-word self-paced reading task including Garden Path sentences. Using this design, the study teases apart different types of working memory tasks, focally the backwards digit span task and the reading span task. This distinction was made as it has been hypothesised that the correlations between working memory capacity and linguistic processing result in part from the working memory tests involving a lot of linguistic material themselves. By extension a potential source for the correlation is that both the memory tasks and reading tasks find common ground in requiring linguistic experience rather than directly relating to working memory. Farmer et al. (2017) therefore focussed on the linguistic based reading span task, involving the keeping in memory of words and reading of sentences simultaneously and the backwards digit span task, similar in nature but without linguistic material. The study found that the magnitude of Garden Path effect, and therewith the reading time increase, positively correlated with a parser's reading span score, meaning it took a participant longer to read a Garden Path sentence when that participant had a higher reading span score. The study also found a negative correlation between the backwards digit span task and the magnitude of the Garden Path effect. The difference between both memory tasks can be explained by the reading span score correlating with measures of linguistic experience while the backwards digit span task score did not (Farmer et al., 2017).

Similarly, further evidence for the *Dedicated Resources* account comes from a study by.

Traxler et al. (2012). The study analysed eye-tracking data and whether reading speed and working memory were a more significant predictor. Their analysis resulted in favour of reading speed, suggesting that the sentence reading process is not primarily modulated by working memory. Van Dyke et al. (2014) investigated comprehension failure as a function of 24 different measures of individual skill, including working memory, receptive vocabulary and IQ. Their analyses suggest that working memory capacity is not a strong predictive factor for performance in reading tasks. They hypothesise that the greatest hurdle for the parsers stems from interference during the retrieval process which does not directly relate to working memory but rather correlates most strongly with receptive vocabulary. Yet, Van Dyke et al. (2014) note a strong correlation between working memory capacity and IQ, which was a significant predictor for performance. This correlation could possibly reconcile the numerous other findings hailing working memory capacity as a major factor in language processing.

1.5 Good Enough Parsing

The potential role of working memory in ambiguity resolution is also a crucial element of the Good Enough parsing hypothesis (Blache, 2016; Christianson et al., 2001, 2006a; Christianson, 2016; Ferreira et al., 2001; Ferreira & Patson, 2007; Traxler, 2014). The Good Enough parsing approach entails that a parser will not always perform a complete analysis of the sentence to be processed but focus on information thought to be minimally necessary to interpret the sentence. Evidence for this claim stems from multiple sources such as idioms, passive constructions, ambiguities, semantic anomalies and Garden Path sentences. Vespignani et al. (2009) investigated idiom recognition, and found that idioms are generally recognised on the third word, and are no longer parsed afterwards. Further, processing of the idiom was facilitated in relation to its prominence, indicating that the parser is subconsciously relying on statistical evidence. Ferreira (2003) auditorily presented participants with non-canonical passive constructions and proceeded to ask them about the agent of the sentence. Ferreira (2003) discovered that the thematic role was frequently being reversed to the canonical order when participants replied hinting at participants not fully parsing the sentence.

Traxler et al. (1998) tested whether there was a difference in reading time between sentences with ambiguous and unambiguous adjunct attachment options. The experiment resulted in the fully ambiguous sentences being read quicker than the unambiguous ones. Ferreira et al. (2002) interprets these results in favour of Good Enough parsing, attributing the difference in reading time to the assumption that the parsers do not fully specify ambiguous sentences, and proceed without resolving it.

Another supporting piece of evidence in favour of the Good Enough parsing approach comes from parsers seemingly ignoring semantic anomalies. Ferreira et al. (2002, 2007) point out that when presented with a question as *"How many of each type of animal did Moses take on the ark?"* parsers tend to provide the answer *"2"*, completely ignoring the fact that Moses canonically never had an ark, but according to the popular story it was rather Noah who took the animals on the ark.

Evidence for the Good Enough parsing approach in relation to Garden Path sentences comes from the follow-up questions probing the relation between the verb and NPs in the sentence such as in (5.). Instead of fully reanalysing the sentence when encountering diverging material, Christianson et al. (2001) found that a parser is likely to not fully abandon the initial interpretation as evidenced by the incorrect answers to the follow-up comprehension questions. (5.) represents an example of the Garden Path stimuli used by Christianson et al. (2001) and the follow-up question.

(5) *Garden Path sentence and follow-up question (Christianson et al., 2001)*

GP: While Anna dressed the baby that was cute and cuddly played in the crib.

Q: Did Anna dress the baby?

As such, participants had the tendency to answer the question in (5) with *yes*. The authors attribute this to lingering interpretations of the thematic roles of the first reading. Good Enough parsing finds a potential source of this lingering interpretation in the Garden Path model . In comparison to other language processing models, the Garden Path model suggests a surprisingly simple and straight up approach resulting from the principles laid out in (1, 2, 3) (Frazier & Fodor, 1978; Kimball, 1973; Vincenzi, 1991), which corresponds with the central idea of Good Enough parsing, suggesting that a parser will not always maximally interpret an utterance, but can be satisfied with a "shallow interpretation". Drawing the parallel between Good Enough parsing and Garden pathing, the resulting misinterpretation could result as a consequence of an incomplete reanalysis (Christianson et al., 2001, 2006). Ferreira & Patson (2007) note that the lingering misinterpretation under the Good Enough parsing approach almost never occurs when the Garden Path stimuli are separated by a grammatically placed comma, lending support to the *Implicit Prosody Hypothesis* as it implies a more complete disambiguation process following a punctuation marker. Finally, relating the study by Kim & Christianson (2017) to Good Enough parsing, it can be hypothesised that response accuracy to the Good Enough parsing questions decreases in a negative relation to working memory task performance. As such the Good Enough parsing approach suggests a parsing model in which Garden Path constructions, the disambiguating nature of punctuation and the role of working memory can be tested.

2. Current Proposal

Extending on the experimental design in Scheifer (2021) the present study investigated English Garden Path sentences with the goal to provide a better understanding of Garden Path effects and lingering interpretations as predicted by Good Enough parsing. The focal point of the study was the disambiguation process in Late Closure Garden Path structures and, more specifically, whether the disambiguation process is impacted by the presence or absence of a grammatically positioned comma, a parsers working memory capacity and the level of delay between the onset of the comprehension question, checking for a successful disambiguation, and the end of the preceding sentence.

To relate these concepts, the study presented participants with a word-by-word self-paced reading task, designed around the stimuli used in Scheifer (2021). These stimuli were manipulated following the supervisor's feedback, and new stimuli were sourced from Gráf (2014) and Galkina (2015) to expand the paradigm. Given the different goals of the studies, only the typical Garden Path- and unambiguous control sentences were maintained. These two conditions then each had a grammatically correct comma placed within, resulting in 4 different conditions, visible in (6).

(6) *Experimental stimuli*

(a.) *No Punctuation Garden Path*
When the man knitted the socks fell to the floor.

(b.) *Punctuated Garden Path*
When the man knitted, the socks fell to the floor.

(c.) *Punctuated, Unambiguous control*
When the man knitted the socks, the keys fell to the floor.

(d.) *No Punctuation, Unambiguous control*
When the man knitted the socks the keys fell to the floor.

The first topic under scrutiny, which the study hinged on, was the presence of Garden Path effects in sentences like 6a, b., when comparing them to their disambiguous counterparts, 6c, d. As Good Enough parsing assumes that the lingering interpretations are the result of a reanalysis, the absence of Garden Path effects would be detrimental, as it leaves no source for the ambiguity to arise. To test for the presence of a Garden Path effect, reading times in the disambiguation region, the verb *fell* in 6a, b., were contrasted to those on the verb *fell* in completely disambiguous structures such as 6c, d. If there was a significant increase in reading time for the temporarily ambiguous conditions, the Garden Path approach takes it as evidence of a reanalysis occurring (Frazier & Fodor, 1978).

The question the study set out to answer was whether there was evidence for lingering interpretations as predicted by the Good Enough parsing account following temporary ambiguities (Christianson et al., 2001, 2006a; Ferreira et al., 2001). To check the validity of this prediction, the comprehension questions following the Garden Path conditions were analysed and contrasted to the replies given following unambiguous stimuli. If the number of correct replies to Garden Path stimuli is significantly lower than the number of correct replies to unambiguous structures, this is taken as support for an incomplete reanalysis under the Good Enough parsing approach.

Another hypothesis the study aimed to check using the paradigm in (6) was the *Implicit*

Prosody Hypothesis (Bader, 1998; Fodor, 1998, 2002) under which a comma serves the same, or at least a similar, function in written input than a prosodic break does in speech. As such, it is predicted to have a disambiguating effect and has been hypothesised to reduce or completely eliminate Garden Path effects (Ferreira & Patson, 2007). To test this claim, the Garden Path stimuli with and without commas were contrasted in their reading time as it was predicted that the facilitation effect would be reflected in a lower reading time. Further, the presence of a comma as a disambiguating element was additionally tested by evaluating their correlation with the replies given to the comprehension questions. If commas serve to disambiguate, it was predicted that their presence should increase the number of correct replies given.

A final topic of interest for the present study was the role of working memory on ambiguity resolution. Since even the nature of working memory itself is not yet entirely clear, the present study used two different tasks to evaluate a participant's working memory capacity. The first of these was a backwards digit span task, where participants were presented a series of individual digits before being asked to recall the series in reverse order. The second working memory task was a reading span task, where the subjects were presented with a series of sentences, from which they had to remember the last word of each sentence and recall the set of words at the end of the series. The reasoning behind these two tests being chosen is that it has been hypothesised that the relation between some working memory tasks and reading time experiments stems in part from a participant's language experience rather than the working memory capacity itself (Farmer et al., 2017; Van Dyke et al., 2014; Zhou et al., 2017). To control for this effect, the present study used the reading span task, involving linguistic material and the backwards digit span, devoid of any linguistic material to have a balanced design. The scores from these working memory tests were then tested against the reading times in the disambiguation region of Garden Path structures. Going by Farmer et al. (2017), the prediction was that reading times would increase as a function of reading span score, with participants with better performance experiencing stronger Garden Path effects. In parallel, it was predicted that reading times in the disambiguating region would decrease as a function of the backwards digit span score.

Additionally, the two working memory scores were fitted in a logistic regression model with the correctness to the comprehension questions as outcome. Under the Good Enough parsing approach, it was predicted that a higher working memory score would lead to more incorrect replies, as a high working memory participant would be able to maintain the initial analysis for longer, leading to competition (Kim & Christianson, 2017). In addition, the study introduced two levels of delay between the onset of the question and the end of the preceding sentences, hypothesising that at a longer level of delay the two interpretations would compete for longer and response accuracy drop.

In summary then, the present study aimed to address the following 4 questions; is there evidence of a Garden Path effect in temporarily ambiguous structures reflected by an increased reading time in the disambiguating region? Is there evidence for lingering interpretation such as predicted under the Good Enough parsing approach, measurable in the number of correct and incorrect responses given to comprehension questions following Garden Path structures? Does orthographic punctuation serve as a disambiguating element, decreasing reading time in the disambiguating region and improving response accuracy following ambiguities, similar to prosodic breaks under the *Implicit Prosody Hypothesis?* In which manner does working memory capacity interact with ambiguity resolution, measured by the reading time in the disambiguating region and response accuracy to follow-up questions to Garden Path sentences at different levels of delay?

In line with the previous thesis, the entire experiment was carried out using the online experiment builder Gorilla.

3. Methodology

3.1 Participants

A total of 58 participants (mean age = 37.17, SD = 14.69), all native speakers of English, took part in the experiment. Participants indicated to come from a wide range of countries, including the United States, the UK, Australia and South Africa. To keep the experiment as close to the original as possible, the same exclusion criteria were utilised for the self-paced reading experiment: participants whose reading time data deviates by more than 2.5SDs will be excluded, as well as when their response accuracy to the follow-up questions in the filler conditions falls below 85%. For the Backwards Digit Span task, participants were excluded if their task duration deviated further than 3SDs from the mean. For the Reading Span task, participants were only excluded if they failed to provide answers in the proper format. Finally, if a participant indicated that they suffer from a condition which could impact reading comprehension, such as dyslexia or ADHD, their data was excluded from all tasks as the data they provide risks being too noisy.

Recruitment of participants took place through *Prolific*, pre-screening for only native speakers of English.

3.2 Stimuli

3.2.1 Working memory measures

3.2.1.1 Working memory measures - Backwards Digit Span Task

The stimuli of the backwards digit span task were fully randomised digits ranging from 1 to 9 arranged in strings of varying sizes. In total, there were 7 sets of two equally long strings of digits, with string size ranging from 2 to 8, resulting in 14 different digit strings total. Digits were free to repeat within a string as long as it was not in immediate succession and set size increased incrementally, starting off with the two sets of digit string length 2, and finishing on those with digit string length 8.

Set Size	Stimuli	Answer
2	1 5	5 1
3	4 5 3	3 5 4
4	8 6 4 8	8 4 6 8
5	1 4 9 5 8	8 5 9 4 1
6	8 4 6 3 5 8	8 5 3 6 4 8
7	3 5 9 4 1 8 2	2 8 1 4 9 5 3
8	4 1 7 6 3 5 3 1	1 3 5 3 6 7 1 4

Table 1: Example stimuli sets from the Backwards Digit Task

3.2.1.2 Working memory measures - Reading Span Task

The stimuli of the reading span task stem from the non-automated Eprime 1.0 script found on https://englelab.gatech.edu/, who offers free memory tasks for research purposes. In accordance with Oswald et al. (2015), for a shortened version of the test, a total of 30 sentences, both semantically felicitous or fallacious were selected.

	Sentence	Semantically felicitous?	
Stimuli 1	Before the test started, he wished us good **luck**.	Yes	
Stimuli 2	The table sprints at the **wall**.	No	
Word Recall			**luck wall**

Table 2: Example sentence set from the Reading Span Task

3.2.2 Self-paced reading task

As alluded to above, the stimuli for the self-paced reading task were adapted from Scheifer (2021), Gráf (2014) and Galkina (2015) (Appendix A). These were then further manipulated within the framework of the present study. As such, the paradigm featured prototypical non punctuated garden path sentences (7a.), punctuated Garden Path sentences (7b.), punctuated, unambiguous control sentences (7c.) and non punctuated unambiguous control sentences (7d.). The present study maintained two out of the four conditions from Scheifer (2021), exchanging the remaining two for conditions relevant to the current study. A distinction was the introduction of the presentation delay between the onset of the comprehension question and the end of the preceding sentence to check for the influence of working memory on comprehension accuracy after different time intervals. (Appendix A)

(7) Experimental conditions

(a.) *No Punctuation Garden Path*
When the man knitted the socks fell to the floor.

(b.) *Punctuated Garden Path*
When the man knitted, the socks fell to the floor.

(c.) *Punctuated, Unambiguous control*
When the man knitted the socks, the keys fell to the floor.

(d.) *No Punctuation, Unambiguous control*
When the man knitted the socks the keys fell to the floor.

The *No Punctuation Garden Path* condition (7a.) consisted of classical examples of Garden Path constructions. These sentences are designed to mislead the reader into thinking that the second NP, *the socks,* act as the direct object of the preceding verb *knitted*. Upon reaching the second verb, *fell*, participants were predicted to realise the error in their initial analysis and start the reanalysis process. It is predicted by the Garden Path model that this reanalysis will be reflected in an increased total reading time, resulting from the regressive analysis. To allow for this misinterpretation by the participant, the first verb in this condition was always potentially transitive or intransitive, while the second NP consisted of a likely complement. The stimuli from this condition are further misleading as they omit any mid-sentence punctuation and thereby fail to alert the reader of the upcoming syntactic structure. Garden Path stimuli deemed fit by the examiners of Scheifer (2021) were maintained and those unfit replaced by sentences tested in Gráf (2014) and Galkina (2015) or modified following the supervisor's advice.

The *Punctuated Garden Path* condition (7b.) consisted of the same sentences as the *No Punctuation Garden Path* condition. However instead of stranding the participant without punctuation, a grammatically correct comma was inserted after the first verb, *knitted,* predicting that it will facilitate a correct first analysis and by consequence reduce the likelihood of a reanalysis

(Drury et al., 2016; Steinhauer, 2003). A potential effect which might be observed is a wrap-up effect at the NP following the comma, *the socks* in (7b.), stemming from the participant parsing the closure indicated by the comma (Drury et al., 2016).

The *Punctuated Unambiguous control* (7c.) was made up of stimuli which resembled those from the *No Punctuation Garden Path* condition but differ in two crucial elements. The first difference was the presence of a grammatically correct comma following the second NP *the socks*, indicating a phrase boundary and therewith aiding disambiguation. The second difference came from the presence of a third NP, *the keys*, serving as an unambiguous subject for the matrix clause. As such, the *punctuated unambiguous control* condition was able to serve as a control to verify that the *No Punctuation Garden Path* condition elicited the predicted effects.

The *No Punctuation Unambiguous control* condition (7d.) is the exact same as *Punctuated Unambiguous control* condition apart from the fact that this condition does not have a comma included. Through comparison with the punctuated counterpart, this condition allows to control for the magnitude of wrap-up effects introduced by the commas to *Punctuated Garden Path* condition (7b.).

In general, all the target sentences followed the same structure. The first word of the sentence was a *wh*-element, either *while* or *when*. This was consistently followed by a DP referring to an animate noun. This was followed by a verb which was potentially transitive, but could also be intransitive. This verb served as the source of the temporary ambiguity within each sentence. Following the first verb, a second DP followed in every condition. This DP was selected to be a plausible direct object of the preceding verb, but in Garden Path conditions (7a., b.) it ended up acting as the subject of the following verb.

In the unambiguous conditions (7c., d.), the second DP ended up being the direct object of the preceding verb, and the following third DP, absent in the Garden Path conditions, acted as the subject of the matrix clause.

The matrix-subject DP was followed by either an adverb relating to the following verb or by a verb followed by a prepositional phrase, depending on what was evaluated to be more natural.

For each target sentence, participants were also faced with a filler construction. The fillers came in different forms, varying in different elements from the target stimuli. For example, they potentially employed proper names and did not always lead with the subclause. The fillers also introduced different *wh*-elements to the experiment as to reduce the risk of participants catching on to the design. The most important difference though was that the fillers were never temporarily ambiguous (8). The full list of fillers can be found in Appendix B.

(8) *Examples of filler constructions*

(a.) The conductor did not know why the train stopped

(b.) When Bert came home, Greta greeted him.

After each sentence, participants were asked to answer a *yes/no* comprehension question (Table 3).

For both control conditions the questions checked whether the participants would competently answer to non-ambiguous constructions. (Table 3)

For the ambiguous target sentences (7a., b.) the questions aimed to check whether the participants would competently answer to non-ambiguous constructions. For the *Punctuated Garden Path* and the *No Punctuation Garden Path* conditions the question served to control for a successful reanalysis and if the first analysis had been completely abandoned or if the reader was satisfied with "shallow parsing" (Blache, 2016; Christianson et al., 2001, 2006; Ferreira et al., 2001; Ferreira & Patson, 2007; Traxler, 2014). To investigate the impact of working memory on ambiguous sentence resolution, the questions were presented with two degrees of delay (500ms, 1500ms) between the participant pressing the spacebar to continue and the onset of the question, with levels of delay distributed evenly across conditions.

With respect to the filler stimuli, the following questions (Table 3) controlled whether the reader was attentive or not during the experiment. Since the sentences that preceded the comprehension questions of fillers were neither potentially ambiguous nor complex for a native speaker of English, the response accuracy on these questions was taken as an exclusion criteria.

Condition	Sentence	Question	Correct answer
No Punctuation Garden Path	When the man knitted the socks fell to the floor.	Did the man knit the socks?	No
Punctuated Garden Path	When the man knitted, the socks fell to the floor.	Did the man knit the socks?	No
Punctuated Unambiguous Control	When the man knitted the socks, the keys fell to the floor.	Did the man knit the socks?	Yes
No Punctuation Unambiguous Control	When the man knitted the socks the keys fell to the floor.	Did the man knit the socks?	Yes
Filler ("no" answer)	When Bert came home, Greta greeted him.	Did Greta come home?	No
Filler ("yes" answer)	The athlete did the exercise while sweating profusely.	Was the athlete sweating while doing the exercise?	Yes

Table 3: Experimental conditions and fillers and their corresponding questions

3.3 Procedure

The consent form, background questionnaire, self-paced reading task, reading span- and backwards digit span tasks were designed in *Gorilla.sc*. As such, participants were able to conduct the experiment without having to come to the laboratory. Recruitment of participants was done exclusively through *Prolific*, a recruitment tool which allows to prescreen for a number of factors, including the participants being native speakers of a language.

Before beginning with the main task, participants were presented with a consent form and subsequently, if consent was given, were asked to confirm whether their native language was English, as required by the recruitment platform *Prolific*. Participants who did not consent or indicated that they were not native speakers of English were rejected from the experiment in accordance with the guidelines on *Prolific* and not considered further. (Appendix D)

After these two preparatory stages, the participants were taken to the self-paced reading experiment. To begin, they were shown an instruction screen, explaining that they were about to begin a word-by-word self-paced reading task. Following the instructions, participants were shown two trial sentences with the goal of familiarising them with the experimental setting, and at the end of them, were presented with the option to return to the instructions in case something seemed unclear. The task consisted of 72 sentences, 36 target sentences (7) and 36 filler sentences (8). Each sentence was followed by a yes-or-no question relating to the preceding sentence. Before each sentence, a fixation cross was being displayed at the centre of the screen for 300ms. After the fixation cross, the sentences were displayed on a word-by-word basis, centred around the same location. Participants had to click “Next” at the bottom of the screen in order to advance to the next word. The end of each sentence was marked by a dot. Clicking “Next” on the final word of a sentence led to a screen reading “Question:" to appear. This screen lasted for either 500ms or 1500ms depending on the level of delay. The task was therefore similar to the self-paced reading task in Scheifer (2021).

After completing the self-paced reading task, participants were taken to the reading span task. As the current experiment involved multiple tasks, shortened versions for the working memory tasks were chosen so as to not burden participants too much.

The reading span task presented participants with semantically felicitous or fallacious sentences, which belonged to sets of incrementally increasing size, ranging from set size 4 (sets with 4 sentences) to set size 6 (sets with 6 sentences). Each set size occurred twice for a total of 30 sentences. The task began by presenting the participant with instructions, explaining that they would have to read sentences and remember the final word of each sentence. Further, after each sentence the participant would be asked to evaluate whether the sentence was semantically felicitous or not. After the instruction screen, participants were given 5 trial sets, ranging from set size 2 to set size 6. Afterwards, the main reading span task began. To proceed to the semantic felicity judgement, participants had to click “Next” at the bottom of their screen. To make a judgement on the preceding sentence, participants had to answer the question: “Was the preceding sentence semantically felicitous?” with the options to click either “yes” or “no”. The semantic acceptability judgements were not considered in any analysis but rather were included to insure that the subjects read the entire preceding sentence and not only focus on the final word. After the participant rated the sentence, a fixation cross appeared at the centre of the screen for 300ms before the next sentence was shown. At the end of a set, three “?” were being displayed for 3000ms, before a screen would ask them to recall the words in correct order and to insert a space between them. Participants could continue to the next set by hitting Enter. A time limit of 12 seconds to read each sentence was imposed to assure that participants would not linger at a sentence trying to commit the preceding words to memory. At the end of the task, participants were given feedback, providing them the number of sets recalled correctly and incorrectly out of the total number of sets asked.

After the reading span task, the participants were taken to the backwards digit span task. The task began again with an instruction screen, informing the participants about the nature of a backwards digit span task and their assignment ahead of them. They were also informed about the upcoming trial set, before being asked to answer it. The trial consisted of two sets of two digits, used to familiarise the subjects with the position of the digits on the screen. Digits appeared in the centre and were preceded by a fixation cross for 300ms and remained for 200ms.

For the main task, similar to the reading span task, the stimuli were also sets of varying sizes, containing a different number of digits. The smallest set consisted of two digits, meaning that at the end of the set, the participant had to repeat two digits in reverse order. The largest set consisted of 8 digits. Each set size was included twice in the main task, resulting in 14 sets total. Presentation of the sets proceeded incrementally, starting at set size 2, and increasing to set size 8 at a rate of 1. When participants reached set size 8 for the first time, the task began anew at size 2. Upon answering to 14 sets, the experiment was concluded. At the end of the task, feedback was provided as to how many sets were recalled correctly and incorrectly in their entirety.

After concluding the self-paced reading and both working memory tasks, participants were taken to a short background questionnaire to assess their linguistic background and indicate whether they suffered from any condition which might impair active reading. The background questionnaire further asked participants to indicate their age and dominant hand. (Appendix C)

After finishing the self-paced reading task, both working memory tasks, and the background questionnaire, participants were taken to an optional form where they could express their thoughts on the experiment and provide feedback if wanted. (Appendix E)

Clicking “Next” on this final screen returned the participant to the recruitment platform *Prolific* and made them eligible to receive their compensation. Participants were awarded with 6£ an hour..

4. Analysis

4.1 Data exclusion

4.1.1 Self-paced Reading Task

Participants whose total reading time deviated more than 2.5 SDs from the group mean were excluded from the analysis of the self-paced reading task. This led to the exclusion of 3 participants. Subsequently, answers to the comprehension questions following filler sentences were evaluated, and participants failing to answer 85% of the basic comprehension questions were also excluded as this was taken as careless execution of the task (Appendix A). This excluded 3 further participants If a participant indicated to suffer from a reading impairing disorder, the data was also not included in the analysis.

Based on these criteria, a total of 6 participants were excluded. Taking into account the participants which had to be excluded due to failure on other tasks, the total number of participants maintained was 37.

4.1.2 Working memory measure - Backwards Digit Span Task

For the backwards digit span task only participants whose total task time was removed further than 3 SDs from the mean were excluded. This led to the exclusion of 1 participant. Some of the participants' responses went against the instructions and included commas, spaces or a combination of both. These responses were trimmed in Excel, removing all punctuation and white space and subsequently included in the analysis. Taking into account the participants which had to be excluded due to failure on other tasks, the total number of participants maintained was 37.

4.1.3 Working memory measure - Reading Span Task

For the reading span task, 17 participants were excluded on the basis that during the word recall, instead of providing the last word of each sentence previously presented, these participants replied by providing the final sentence of each set. This form of reply was interpreted as a participant's failure to understand the task at hand . As a consequence, 41 participants remained for the analysis of the reading span task. Some other participants' responses went against the instructions and included commas, spaces or a combination of both. These responses were trimmed in Excel, removing all punctuation and white space and subsequently included in the analysis.

Taken together, 37 participants completed all three tasks and their data was used for the final analyses and fitting of the models.

4.2 Data analysis

4.2.1 Working memory measures

4.2.1.1 Working memory measures - Backwards Digit Span Task

To score the Backwards Digit Span task, participants received one point for each consecutive digit string which was correctly recalled in its entirety. Failing two trials of the same set resulted in the scoring being halted for that participant. As such, a participant could score between 0 and 14 (Farmer et al., 2017; Raiford et al., 2010). Each participant had their replies evaluated by a Python script assuring consistent scoring across participants.

4.2.1.2 Working memory measures - Reading Span Task

To score the reading span task, each consecutive fully recalled stimulus awarded the participant with a point. Having one out of the two responses correct of the following set size awarded them with half a point. As such, the scores range between 0 and 6 (Farmer et al., 2017). Scoring of the reading span task was done using a Python script.

4.2.2 Self-paced Reading Task

Analysis of the data followed the same critical regions as Scheifer (2021) as far as possible. As such, five regions of interest were designated (Table 4).

Condition	wh - element	Region 1	Region 2	Region 3	Region 4	Region 5
No punctuation GP	When	the man	knitted	the socks	N/A	fell to the floor.
Grammatical punctuation GP	When	the man	knitted,	the socks	N/A	fell to the floor.
Disambiguous no punctuation	When	the man	knitted	the socks	the keys	fell to the floor.
Unambiguous, grammatically punctuated sentence	When	the man	knitted	the socks,	the keys	fell to the floor.

Table 4: Stimuli and critical regions for each condition.

. For the analysis of the self-paced reading task, the sentences were divided into 5 regions. Prior to Region 1, each sentence began with a w*h*-element to introduce the subclause, varying between *when* or *while*. Region 1 consisted of the first DP of each condition, *the man*, in Table 4. This was selected as a region of interest as it represents the subject of the embedded clause, and therewith the subject of the temporarily ambiguous verb.

Region 2 consists of the first verb of each condition, *knitted*, in Table 4. This verb is crucial as it is not the only source of ambiguity, allowing for transitive or non-transitive constructions. Further in the punctuated Garden Path condition, this verb is followed by a comma predicted under the *Implicit Prosody Hypothesis* to disambiguate the Garden Path construction.

Region 3 is made up of the second DP of each condition, *the socks in* Table 4. This DP can either serve as the object of the preceding verb (Region 2), or subject of the matrix clause. In the

Unambiguous Punctuated control condition, this DP is also followed by a comma, which is predicted to further disambiguate this sentence. Combined with Region 4, only present for the disambiguated conditions and consisting of the third DP present in these conditions, it allows to control for potential wrap-up effects which are predicted to surface as increased reading time on the words following a comma (Drury et al., 2016).

Region 5 consists of the final verb and the following material. If an adverbial preceded the verb, this was also included in Region 5. This region was of critical importance for the Garden Path approach, as it is here that the reader is being made aware that the DP they read in Region 3, actually acts as a subject of the matrix clause rather than object of the embedded clause. As such, the No Punctuation Garden Path condition is predicted to have a longer reading time for Region 5 than any of the other conditions.

To check for the presence of Garden Path effects, the reading time of the No Punctuation Garden Path condition will be contrasted in Region 5 to Disambiguous No Punctuation condition.

To check whether the presence of grammatically licit punctuation reduces the magnitude of Garden Pathing, and therefore supports the *Implicit Prosody Hypothesis,* the No Punctuation Garden Path condition will be contrasted to the Punctuated Garden Path condition in Region 5.

To control for wrap-up effects, the No Punctuation Garden Path condition will be contrasted to the Punctuated Garden Path condition in Region 3. In addition, the disambiguous conditions will be compared in Region 4.

To account for the predictions made under the Good Enough parsing approach, the answers to questions following the Garden Path condition will be scored as % correct, and then contrasted to the % correct of replies to the questions following the unambiguous conditions. If the Good Enough parsing approach holds, there should be a significant difference between both sides of the comparison, with the questions following the ambiguous constructions having a lower success rate.

To account for the role of working memory in ambiguity resolution, the working memory scores of each participant will be correlated to their answer to the questions following the ambiguous conditions at each of the two levels of delay. A lower success rate on long delay questions has been interpreted as the lingering interpretation of the first, false, analysis competing with the reanalysed structure (Ferreira & Patson, 2007). This prediction is borne out of the fact that a high reading span score has been found to correlate positively with stronger Garden Path effects (Farmer et al., 2017). The same study also found a negative correlation between the magnitude of Garden Path effects and the scores obtained from a backwards digit span task.

5. Results

5.1 Descriptive statistics

5.1.1 Backwards Digit Span task

Scores on the backwards digit span task could range from 0 to 14. The mean, range and standard deviation for the backwards digit span test are presented in TABLE 5. The data shown in TABLE 5 pertains only to the 37 participants which were included in the final analyses. Figure 1 depicts a histogram representing the distribution of the backwards digit span scores. The most frequent score observed for the backwards digit span task was 9 out of 14 possible points.

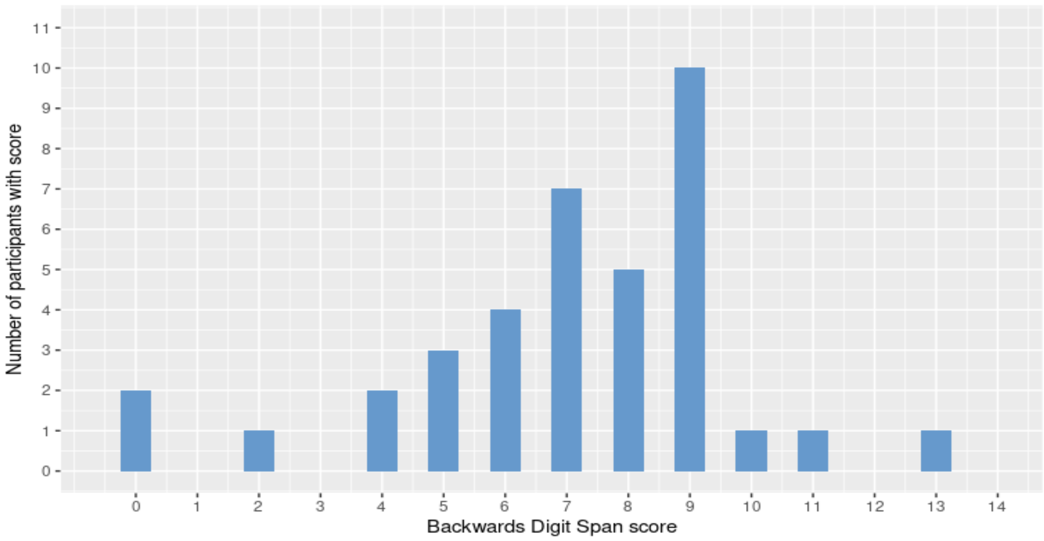

Figure 1: Histogram showing the number of participants achieving a certain score on the backwards digit span task. $N = 37$.

5.1.2 Reading Span task

Scores on the reading span task could range from 0 to 6. The mean, range and standard deviation for the reading span task are presented in Table 5. The data shown in Table 5 represents only the 37 participants which made it through to the final analysis. Figure 2 shows the distribution of reading span scores as recorded in the present study. The most frequent score for the reading span task was 0.5 out of 6 possible points.

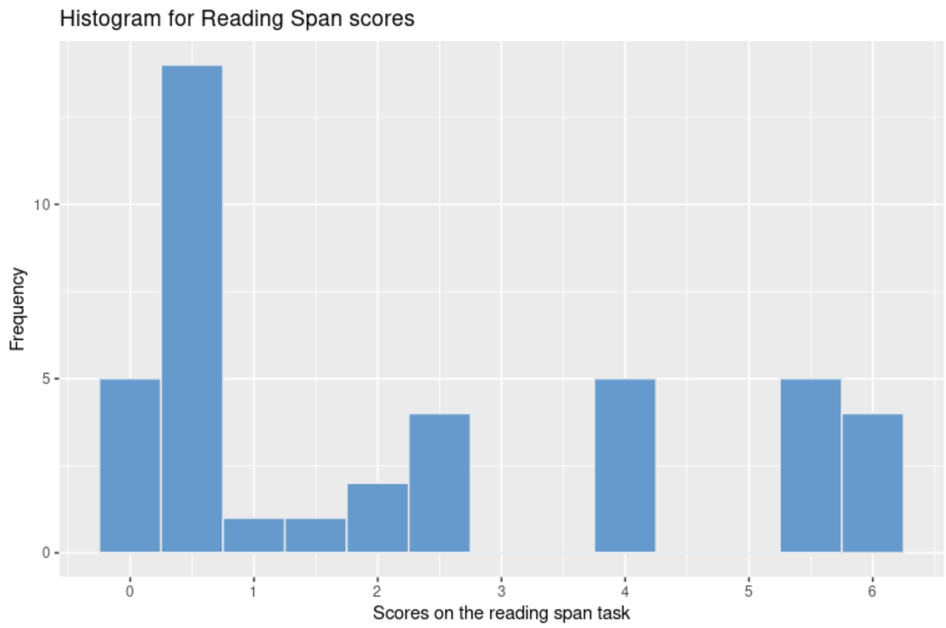

Figure 2: Histogram showing the number of participants achieving a certain score on the reading span task. *N* = 37.

Task	Possible range	Observed range	SD	Mean
Backwards digit span	0-14	0-13	2.67	7.81
Reading span	0-6	0-6	2.25	2.35

Table 5: Descriptive statistics for the backwards digit span and reading span tasks. *N* = 37.

Comparing the data in Table 5 to the results found in Farmer et al. (2017), it is striking that for either memory task, the mean performance is lower for the present study than the referenced study (Reading Span: 2.35 vs. 4.43 | Backwards digit span: 7.81 vs. 9.47). A potential reason for this might be the online nature of the present experiment, rendering the researcher unable to intervene or correct the participant's interpretation of each task. This explanation is further supported by differences in the observed ranges, with the participants in the Farmer et al. (2017) study not once scoring a 0, while in the present study this score has been observed more than once for either test.

5.1.3 Self-paced reading task

5.1.3.1 Total Sentence Duration

Prior to the analysis of the total sentence duration, all sentences with a duration greater to 20'000ms had been removed from the data. Subsequently, boxplots were constructed to identify further outliers (Figure 3). Analyses were limited to the first 8 words for the Garden Path conditions

(7a, b.) and the first 10 for the disambiguous conditions (7c, d.) to amend for differences in stimuli length which originally ranged from 8 to 13 words. Analyses further only include the 37 participants who managed to absolve the three experimental tasks to a sufficient degree, as outlined in 6.1.

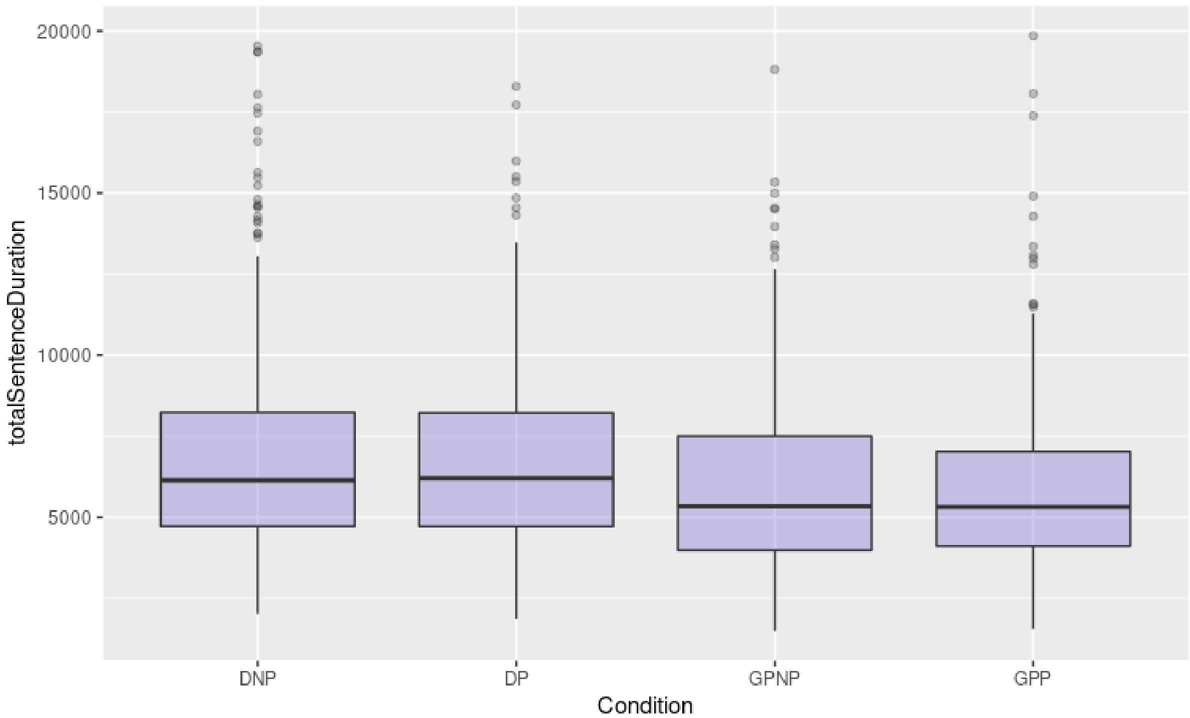

Figure 3: Boxplots for the total sentence duration by condition. DNP = disambiguous, no punctuation, DP = disambiguous punctuated, GPNP = garden path no punctuation, GPP = garden path punctuated

These outliers were removed prior to the descriptive observations (Table 6) and plotting of the sentence duration by condition (Figure 4).

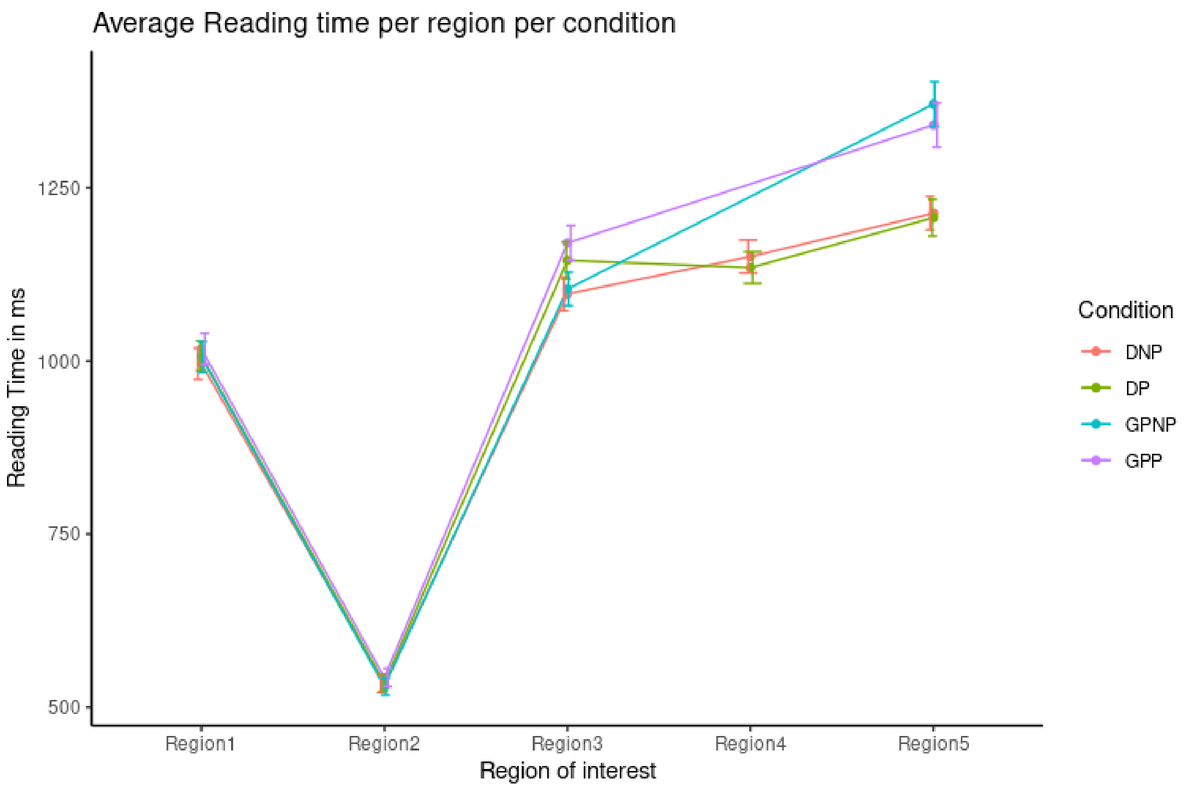

Figure 4: mean reading time in ms per region by condition. DNP = disambiguous, no punctuation, DP = disambiguous punctuated, GPNP = garden path no punctuation, GPP = garden path punctuated

Figure 4 shows the reading times per region by condition, including error bars. Example sentences for each condition with their respective region are repeated in (9). Note that the graph does not include the first word of each sentence, the *wh*-element, and that Region 5 was potentially not the final word of the sentence, merely the disambiguating region.

(9) *Examples of Garden Path and Unambiguous sentences and the respective regions*

Garden Path

(a.) While	the mother	bathed (,)	the baby	happily played.
	Region 1	*Region 2*	*Region 3*	*Region 5*
(b.) While	the mother	bathed	the dog (,)	the baby
	Region 1	*Region 2*	*Region 3*	*Region 4*
	happily played.			
	Region 5			

Taking into account that the Garden Path conditions do not have a Region 4, the largest difference which can visually be observed in Figure 4 is the difference between the unambiguous and Garden Path conditions in Region 5. The dip in Region 2 can be explained as it constitutes the only one-word region and therefore a lower average reading time is expected. All conditions remain fairly close up to Region 5, where a visually significant gap opens.

Condition	Sentence Length (in words)	Mean duration in ms	SD (duration)	Mean in ms (adjusted for sentence length)
Disambiguous no punctuation	10	6964.53	3342.421	5,571.62
Disambiguous punctuated	10	6717.42	2873.89	5,373.94
Garden Path no punctuation	8	5966	2810.56	5966
Garden Path punctuated	8	5839.54	2697.4	5839.54

Table 6: Mean and standard deviation of total sentence duration by condition.

5.1.3.2 Descriptive statistics for Good Enough parsing

To investigate whether the lingering interpretations suspected by the Good Enough parsing approach were still present, the comprehension questions following Garden Path sentences (7a, b.) were compared to those following the disambiguous constructions (7c, d.). Figure 5 shows the distribution of correct or incorrect responses by condition.

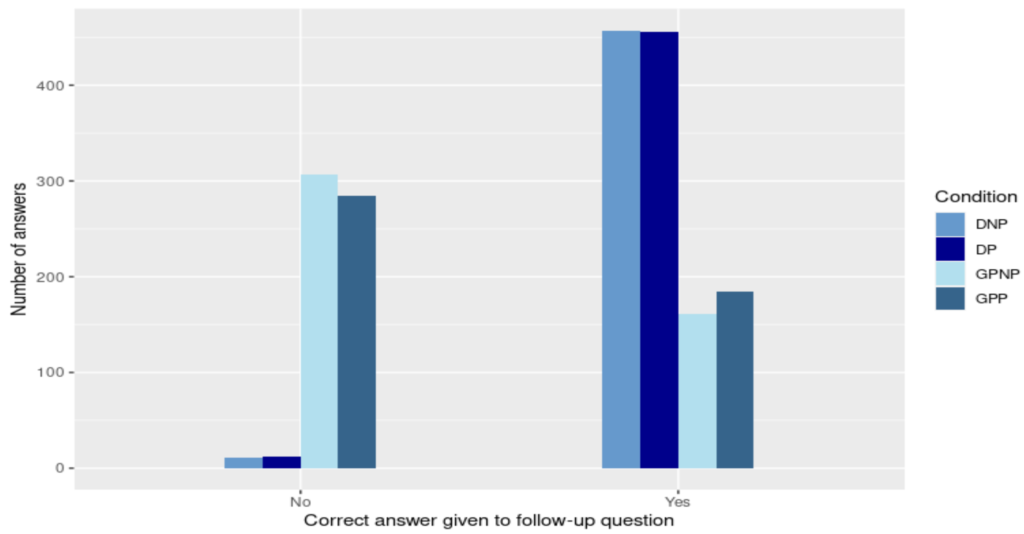

Figure 5: number of correct or incorrect responses to the follow-up question by condition. N = 1872

The data for each condition is made up of 468 observations, adding up to 1872 in total. Immediately, a clear difference between the disambiguous and the Garden Path conditions becomes apparent. Table 7 summarises the number of replies by condition.

Condition	Correct Reply	Incorrect Reply	Total
DNP	457	11	468
DP	456	12	468
GPNP	161	307	468
GPP	184	284	468
	1258	614	1872

Table 7: number of correct or incorrect replies to the follow-up question by condition.

With regards to the Good Enough approach, merging the Garden Path sentences and disambiguous constructions into their respective categories, it results in the following spread (Table 8):

Ambiguity	Correct Reply	Incorrect Reply	Total
Unambiguous	913 (97.54%)	23 (2.46%)	936
Ambiguous	345 (36.86%)	591 (63.14%)	936
	1258	614	1872

Table 8: number of correct or incorrect replies to the follow-up question by ambiguity.

To further investigate the present data, a logistic regression analysis was performed on the data. (Table 13, Table 14)

5.1.3.3 Differences between regions

At first glance, there seems to be a difference between the disambiguous sentences and the Garden Path sentences in terms of mean reading time when adjusting for sentence length with the disambiguous sentences being read quicker than the Garden Path sentences. The impact of punctuation on reading time does appear to be minimal going by the descriptive data. Predictions, observation of means (Appendix F, Figure 4) and the boxplots in Figure 6 point towards Region 5 being the differentiating factor.

While a more detailed analysis of Region 5 will follow, Appendix F shows the mean and standard deviations of each region by condition. Prior to the construction of the boxplots in Figure 6, the reading times for all regions above 4000ms were eliminated. Based on these boxplots, further outliers were then removed before calculating the values in Appendix F. Taking a look at the table in Appendix F, it seems as if the differences before Region 5, the disambiguating region, predicted to be the site of Garden Path effects, seem minimal and that the visually observed difference in Region 5 merits further investigation. A general analysis of Region 5 had been planned a priori.

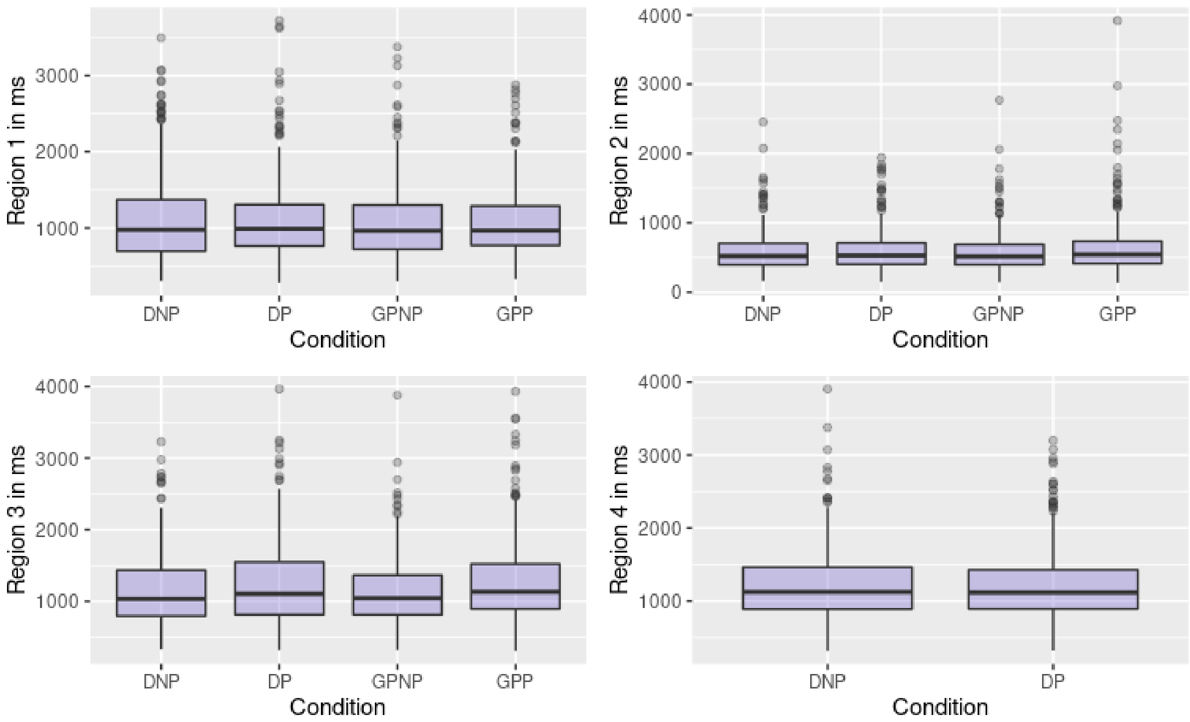

Figure 6: Boxplots showing the distribution of mean reading time (in ms) of the first four regions by condition. DNP = disambiguous, no punctuation, DP = disambiguous punctuated, GPNP = garden path no punctuation, GPP = garden path punctuated

5.1.3.4 Descriptive statistics for Region 5

With Region 5 containing the disambiguating verb suspected to evoke a reanalysis under the Garden Path approach (Frazier & Fodor, 1978), a reading time analysis of this region had been decided on a priori. Following the observation of the descriptive statistics pertaining to the total sentence duration and the intuitively small differences observed between the preceding regions, Region 5 does seem to be the determining region. Again, prior to the construction of the boxplots for Region 5, the region was filtered to exclude all data points above 4000ms. Following this initial filter, the boxplots in Figure 7 were constructed to check for any further outliers.

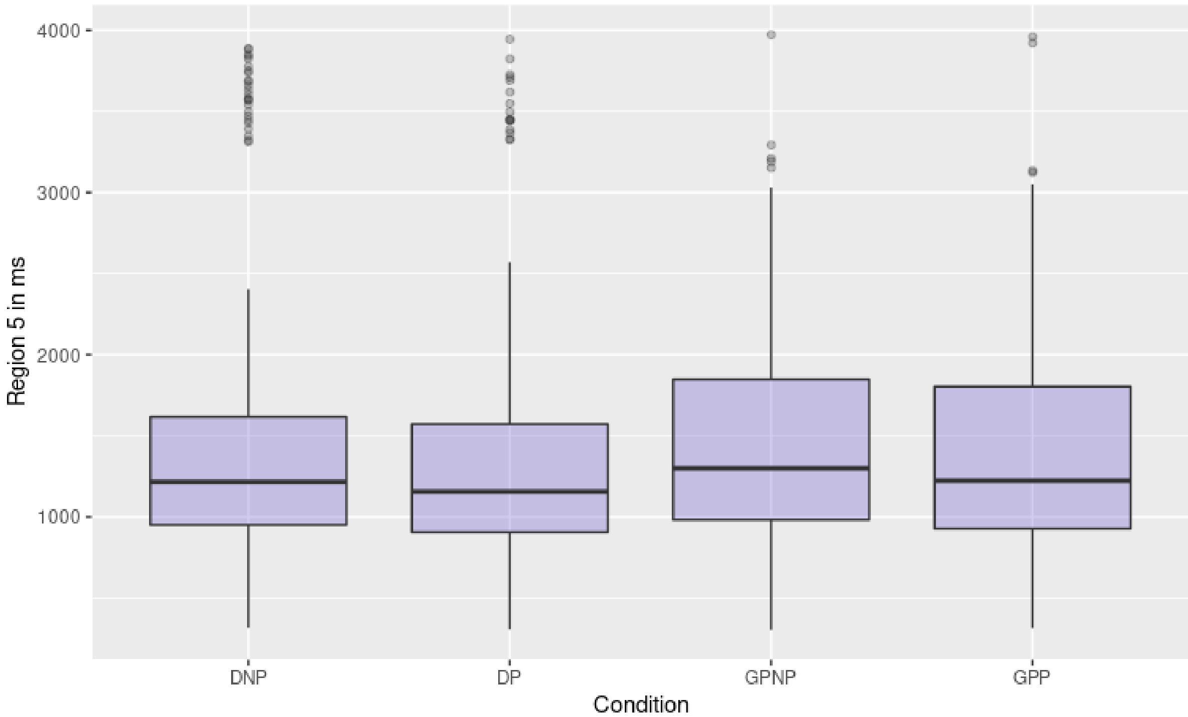

Figure 7: Boxplots showing the distribution of mean reading time (in ms) of Region 5. DNP = disambiguous, no punctuation, DP = disambiguous punctuated, GPNP = garden path no punctuation, GPP = garden path punctuated

After exclusion of these outliers, the values in Table 9, show the mean and standard deviation within Region 5 by condition.

Condition	Mean reading time Region 5 (in ms)	SD Region 5(in ms)
Disambiguous no punctuation	1213.38	443.08
Disambiguous punctuated	1206.88	480.95
Garden Path no punctuation	1370.911	594.28
Garden Path punctuated	1340.66	583.78

Table 9: Mean and standard deviation of Region 5 by condition in ms.

In comparison to the study by Scheifer (2021) which had the equivalents of the disambiguous and Garden Path conditions without punctuation, mean reading times for Region 5 are lower. Scheifer (2021) found a mean reading time of 1848.87ms for the Garden Path condition and 1725.74ms for the unambiguous condition. The difference in reading time in the final region seems rather large between both studies. A possible explanation for this discrepancy is that data filtering for the present study was more rigorous than for Scheifer (2021).

5.2 Reading time analysis of Region 5 - potential Garden Path effect

To assess whether a Garden Path sentence elicits the expected Garden Path effect, linear mixed effect models were fitted with the “lme4” package in R (Bates et al., 2015). Reading times in Region 5 were log transformed prior to the modelling to improve the normality of the distribution (Box & Cox, 1964). The predictors *Backwards digit span score, Reading span score* and *Condition* were tried as fixed effects. For the random effects, *Participants* and *Items* were used for both their intercepts, and the slopes in relation to the fixed effects. The best fitting model, best accounting for the log transformed reading time in Region 5, consists of the fixed effect of *Condition* and the random intercept for *Participants*. *Condition* was the first fixed effect to be added to the null model and significantly improved the fit. The fixed effects for the *backwards digit span* and *reading span scores* were added afterwards but did not further improve the model fit. Next, the random intercept of *Participant* was added, further improving the model fit. The random slope for *Participant* by *Condition* did not improve the model. The model is summarised in Table 10.

	Estimated ß	*SE*	*t*
Intercept	7.05	0.06	113.769
Condition DP	-0.02	0.02	-0.886
Condition GPNP	0.11	0.02	5.02
Condition GPP	0.07	0.02	3.27

Table 10: Regression coefficients and test statistics from the linear mixed effect model for Condition in Region 5. DNP = disambiguous, no punctuation, DP = disambiguous punctuated, GPNP = garden path no punctuation, GPP = garden path punctuated

The positive effects for the conditions GPNP (ß = 0.11, *SE* = 0.02, t = 5.02) and GPP (ß = 0.07, *SE* = 0.02, t = 3.27) are indicative of the log transformed reading time of Region 5 increasing in these two conditions when compared to the no punctuation disambiguous condition. This is indicative of the Garden Path effect predicted by the Garden Path approach (Frazier & Fodor, 1978) caused by a reanalysis when encountering the disambiguating verb in Region 5. The effect seems stronger for the Garden Path condition with no punctuation (7a.) than for the punctuated Garden Path sentences, hinting at the possibility of the comma helping to disambiguate the sentence structure, falling in line with the *Implicit Prosody Hypothesis.* Given the results of the model in Table 10, the decision was made to do a pairwise post-hoc analysis to check for the interaction between different levels of condition and their impact on the log-transformed reading time in Region 5. To account for family-wise error, the Holm-Bonferroni method was used to adjust the p-values. The post-hoc pairwise comparison was performed in R, using the “phia” package. The results of this pairwise comparison can be found in Table 11.

Interaction	Value	X^2	*p*
DNP : DP	0.02	0.78	0.38
DNP : GPNP	-0.11	25.2	0.000003*
DNP : GPP	-0.07	10.69	0.003*
DP : GPNP	-0.13	35.05	0.00000002*
DP : GPP	-0.09	17.43	0.0001*
GPNP : GPP	0.04	3.29	0.14

Table 11: Holm-Bonferroni adjusted pairwise comparison between Condition with the outcome log-transformed reading time in Region 5. DNP = disambiguous, no punctuation, DP = disambiguous punctuated, GPNP = garden path no punctuation, GPP = garden path punctuated. * indicates significance at the 0.05 alpha level.

From the pairwise comparisons it becomes evident that ambiguity is the key factor to predicting variation in reading times. The two interactions within the ambiguous and unambiguous conditions respectively, DNP:DP ($X^2 = 0.78$, $p = 0.38$), GPNP:GPP ($X^2 = 3.29$, $p = 0.14$), are statistically not significant. The conditions in these two interactions contrasted on whether they were punctuated or not. Taking this into account, it cannot be claimed that punctuation had a significant impact on the log-transformed reading times in Region 5. Moving to the significant interaction effects, when comparing the disambiguous condition without punctuation to the non-punctuated Garden Path condition ($X^2 = 25{,}2$ $p < 0.01$) and the punctuated Garden Path condition ($X^2 = 10.69$, $p = 0.003$), it can be seen that effect is larger for the comparison with the non-punctuated Garden Path indicating that the latter takes longer to read.

Comparing the punctuated disambiguous condition to the non-punctuated Garden Path condition ($X^2 = 35{,}05$ $p < 0.01$) and the punctuated Garden Path condition ($X^2 = 17.43$, $p < 0.01$), it can be seen that effect is larger for the comparison with the non-punctuated Garden Path indicating that the latter takes longer to read.

In general, both Garden Path conditions make a large difference in terms of log-transformed reading times in the disambiguating region and can be regarded as significant predictors.

Taking into account the entirety of Table 11, it can be stated that descriptively the non-punctuated Garden Path takes the biggest toll on the log-transformed reading time of Region 5. On the other hand, the punctuated unambiguous condition has the shortest reading time for the same region. These differences are however not statistically significant and can therefore not be taken in favour of the *Implicit Prosody Hypothesis* (Bader, 1998; Fodor, 1998, 2002).

5.3 Good Enough Parsing analysis - evaluation of answers to comprehension questions

To check whether the response to the Good Enough parsing questions fluctuates as a function of a number of predictors, the data were fit into a logistic regression model. Random effects were not considered. Before trying to fit the fixed effects, the categorical predictors Condition and delay were included in proportion tables, plotted against the outcome of whether the question was answered correctly or not (Table 12). The continuous predictors were plotted in proportion graphs, showing the expected change in the correctness to a question as a result of either working memory score (Figure 8).

Question Correct	Condition				delay	
	DNP	DP	GPNP	GPP	long	short
No	0.02	0.02	0.5	0.46	0.51	0.49
Yes	0.35	0.3	0.14	0.16	0.49	051

Table 12: proportions comparing the categorical predictors to QuestionCorrect.

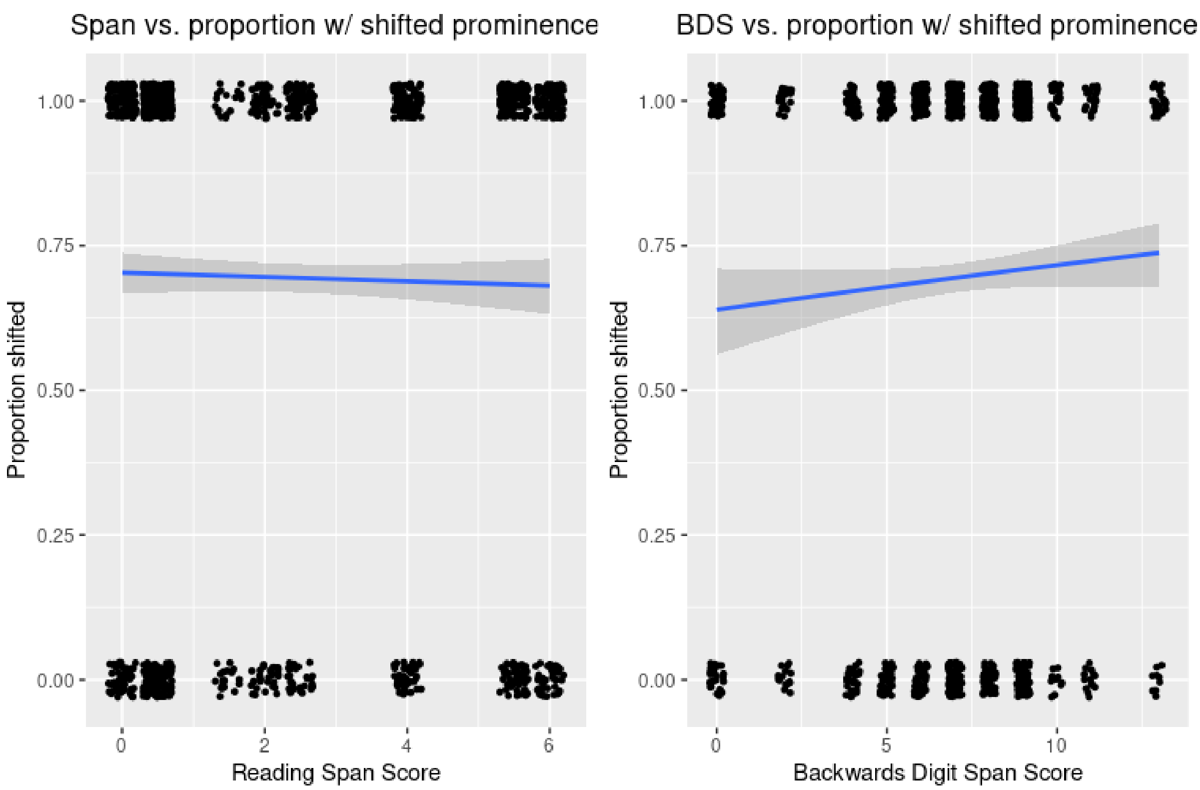

Figure 8: Working memory scores and their predicted influence on question correctness.

From Figure 8 it seems as if the number of questions answered correctly increases as a function of backwards digit span score, while there seems to be a marginal decrease in response to a higher reading span score.
Table 12 shows that a clear effect of *Condition* is to be expected. On the other hand, the effect of delay on the question seems negligible. Resulting from the inspection of the proportion tables and graphs, the model was built bottom-up, trying to fit the four fixed effects of *Condition,, backwards digit span score* and *reading span score, delay* in that order.
Adding the four fixed effects as specified above, checking between each step whether the model fit significantly improved, leads to the final model summarised in Table 13. The final model consists of the fixed effects of *Condition, backwards digit span score* and *delay*. Reading span score did not improve the model fit, nor did the interactions between *Condition* and *delay* or *delay* and *backwards digit span score*. Cox and Snell's, as well as, Nagelkerke's R^2 were calculated using the "DescTools" library for R. Hosmer and Lemeshow R^2 was calculated using the "glmtoolbox" package for R.

	B (SE)	95% confidence interval for odds ratio		
		Lower	Odds Ratio	Upper
Included				
Constant	3.86	20.29	47.41	131.28
Condition DP	- 0.34 (0.55)	0.23	0.71	2.06
Condition GPNP	- 4.58 (0.43)	0.004	0.01	0.02
Condition GPP	- 4.34 (0.43)	0.005	0.01	0.03
BDS score	0.06 (0.03)	1	1.05	1.12
delay short	- 0.37 (0.16)	0.5	0.69	0.93

Table 13: logistic regression summary for the replies to the comprehension questions.
Rounded to the second decimal. R^2 Hosmer-Lemeshow: 0.95, R^2 Cox-Snell = 0.37, R^2 Nagelkerke = 0.52, X^2 = 614.02.
DNP = disambiguous, no punctuation, DP = disambiguous punctuated, GPNP = garden path no punctuation, GPP = garden path punctuated, BDS score = backwards digit span score

Going by the data in Table 13, it can be seen that the Garden Path conditions both reduce the likelihood of receiving a correct response when compared to the non-punctuated unambiguous condition. The effect seems to be slightly larger for the non-punctuated Garden Path condition (*OR* = 0.01, B = - 4.58, *SE* = 0.43) than for the punctuated Garden Path condition (*OR* = 0.01, B = - 4.34, *SE* = 0.43), and for either of them the odds of getting a correct reply are about 100 times less likely than for the non-punctuated unambiguous condition.

The punctuated disambiguous condition (*OR* = 0.71, B = - 0.34, *SE* = 0.55) also seems to lead to a lower number of correct responses when compared to the non-punctuated counterpart, but the effect is much smaller.

Another observation that can be drawn directly from Table 13 is that the short delay of 500ms (*OR* = 0.69, B = - 0.37 , *SE* = 0.16) has received fewer correct responses than the long delay (1500ms).

Counter to the prediction, the lack of a better fit taking into account either working memory score indicates that these did not significantly impact the number of correct answers supplied by a participant, and working memory seems to have played a lesser role, if any, when it came to ambiguity resolution.

Backwards digit span score (*OR* = 1.05, B = 0.06 , *SE* = 0.03) as a fixed effect improved the model fit and positively correlated with the frequency of correct answers indicating that the number of correct replies increases the higher a participant scored on the backwards digit span task.

Since it could be established that the Garden Path conditions are a major contributor to the outcome "QuestionCorrect", as a post-hoc analysis,it was decided to remove the disambiguous conditions from the model and check whether the factor of condition still improves the fit. If it does, this will be taken as evidence for the impact of punctuation in the processing of temporarily ambiguous sentences. This would further allow to check the working memory scores for only the temporarily ambiguous conditions, those they are predicted to make a difference on.

Once again the fixed effects of *Condition*, this time with 2 levels, *backwards digit span score, reading span score* and *delay* were considered. Based on the model in Table 13, *Condition* was the first fixed effect to be added, then *backwards digit span score*, followed by *delay* and finally *reading span score*. Random effects were not considered. Adding *Condition* as a predictor did not lead to a significant improvement over the null model, indicating that punctuation was not impactful when it came to answering the comprehension questions.

Adding *backwards digit span score* as fixed effect does lead to an improvement over the null model. *Delay* as fixed effect further improved the fit. The addition of reading span score does not significantly improve the model further. Neither does the interaction between delay and backwards digit span score. Therefore, the final regression model for the post-hoc analysis of the Good Enough parsing questions for only the ambiguous conditions can be found in Table 14.

	B (SE)	95% confidence interval for odds ratio		
		Lower	Odds Ratio	Upper
Included				
Constant	- 0.61 (0.24)	0.33	0.54	0.87
BDS score	0.06 (0.03)	1	1.06	1.13
delay short	- 0.39 (0.16)	0.5	0.68	0.93

Table 14: post-hoc logistic regression summary for the replies to the comprehension questions. BDS score = backwards digit span score Rounded to the second decimal. R^2 Cox-Snell = 0.02, R^2 Nagelkerke = 0.02, X^2 = 10.01.

From Table 14 it can be observed that a short delay (*OR* = 0.68, B = - 0.39, *SE* = 0.16) reduces the likelihood for a correct response in ambiguous conditions when compared to a long delay. Similarly to the model in Table 13, backwards digit span score (*OR* = 1.06, B = 0.06, *SE* = 0.03) improves response accuracy.

6. Discussion

The study set out to investigate the impact of a number of predictors on the reading time of Garden Path ambiguities and their ambiguity resolution in light of the Good Enough parsing approach (Blache, 2016; Christianson et al., 2001, 2006; Ferreira et al., 2001; Ferreira & Patson, 2007; Traxler, 2014). Two of the predictors in question relating to reading times were working memory scores, as assessed by shortened versions of the traditional reading- and backwards digit span tasks in an online format. Farmer et al. (2017) found a negative correlation between the backwards digit span performance and reading times in Garden Path paradigms, while finding a positive correlation between reading span task performance and the same Garden Path paradigms. Another factor under consideration was the presence of a potentially disambiguating comma separating the matrix and relative clause from each other, which under the *Implicit Prosody Hypothesis* (Bader, 1998; Fodor, 1998, 2002), is hypothesised to be able to serve as a disambiguating element similar to a prosodic break. By extension, the presence of a comma was predicted to accelerate reading times in Region 5 when comparing an non-punctuated Garden Path sentence to a punctuated one (Steinhauer, 2003, 2014; Steinhauer et al., 1999; Steinhauer & Friederici, 2001).

The predictors relating directly to the Good Enough parsing approach were the same working memory task scores mentioned above, as well as the correctness of the answer provided by the participants following a Garden Path construction as opposed to following a regular, unambiguous sentence. Further, the comprehension questions were introduced with two different levels of delay, 500ms or 1500ms and the presence of punctuation was also considered.. Under Good Enough parsing, the presence of a Garden Path effect has been shown to be detrimental to the success rate when answering a comprehension question of the type “Did the husband knit the socks?” (Table 3) (Blache, 2016; Christianson et al., 2001, 2006; Ferreira et al., 2001; Ferreira & Patson, 2007; Traxler, 2014). If participants had a lower success rate answering the follow-up questions after a longer delay, this decrease in success could be interpreted as a consequence of the lingering first interpretation competing with the reanalysis and therefore inhibiting the processing procedure (Ferreira & Patson, 2007). This decrease in success was predicted to be further correlated with working memory performance, with high performance participants retaining the misanalysed form for longer, leading to stronger competition during the delay period. The presence of a lingering interpretation was also predicted to be mediated by the presence or absence of a comma in the target structure (Ferreira & Patson, 2007).

6.1 Limitations of the study

The present study has a number of limitations which should be taken into account before proceeding to the discussion of the results. A principal caveat in this study’s experiment are the working memory tasks. For both the tasks, the performance was descriptively lower than the performance in the reference study by Farmer et al. (2017). A number of possible explanations could be responsible for this discrepancy. To start, the present study was conducted entirely online through the experiment builder Gorilla. By consequence, supervision of the study subjects was not possible and outside of rigorous data filtering, it is impossible for the researcher to control for the participants’ dedication and attentiveness to the study. It was also not possible for participants to ask for clarifications or for the researcher to intervene in case subjects clearly misinterpreted the task they were given. This lack of a controlled environment led to the exclusion of multiple participants from the reading span task, and subsequently the full analyses, as they failed to provide the appropriate answers. A potential, partial solution for this issue could be clearer instructions and trials including feedback before the task. The trials in the present study did not provide the participant with feedback or the correct response afterwards. Including either of these solutions should mitigate the problem of participants not understanding the task at hand. The acquisition of new participants was for temporal and budgetary reasons unfortunately impossible. Further, it is technically possible for participants to cheat on the backwards digit span task by entering the digits in order of presentation and tapping the arrow key to the left to continue writing in front of the digit they just entered. If a participant came up with this strategy before or during the task, it was impossible to intervene or to detect afterwards. This flaw of the online study design seems impossible to circumvent short of logging each of the participants’ keystrokes.

Finally with regards to both working memory tasks, the versions administered were shortened versions of the widely normed and accepted tasks, and for the reading span task specifically, the selection of stimuli, albeit consulting with a native speaker, would have preferably undergone a norming study and each word of the stimuli controlled for frequency. As such they further deviate from the tasks typically administered, including those employed in Farmer et al. (2017).

6.2 Reading Time for Garden Path ambiguities

The Garden Path approach suggests that temporarily ambiguous sentences will have a higher processing time, reflected in a higher reading time for visual stimuli, in the disambiguating region. This increased processing difficulty is assumed to come from the parser sticking with a first analysis, guided by the three principles of Late Closure, Minimal Attachment and the Minimal Chain Principle (1, 2, 3), but then having to revise the entire structure when the disambiguating material comes in. The disambiguating material in the present study was introduced in Region 5. Therefore, if the study found increased reading times in Region 5, this would be indicative of a Garden Path effect.
The results for reading time in Region 5, representing the disambiguating region for the Garden Path stimuli, from Table 9 and Table 10, show a clear trend in increased reading time for the Garden Path stimuli (mean = 1355.79ms) when compared to the unambiguous ones (mean = 1210.13ms) (Figure 4). As such, the presence of Garden Path structure had a significant impact on the reading time of Region 5. The post-hoc analysis modelling ambiguity as a fixed predictor (Table 11) further confirms these findings. From Table 11 it can be taken that, when compared to either ambiguous condition, the disambiguous conditions (ß = -0.1, *SE* = 0.02, *t* = -6.46) take significantly less long to read than their ambiguous counterparts.

Taking into account that the model presented in Table 10 was the best possible fit for the present data, and that none of the post-hoc models had their fit improved by either working memory score, working memory capacity cannot be taken as a significant predictor for log-transformed reading times in a disambiguating region. A potential reason for this result is the deviance in results between the working memory tasks in this study and those in the study by Farmer et al. (2017) which found the correlations between these two tests and reading time in Garden Path ambiguity resolution. As already hinted at in the results section, the participant in this study performed worse than the participants in the study by Farmer et al. (2017) in either test. These lower scores could possibly be explained by the fact the different versions of the tasks were administered, with the tasks in the present study being varying in size, and having undergone less rigorous norming. A different potential explanation comes from studies relating working memory to age. Klencklen et al. (2017), among many others, suggest that working memory declines as a function of age across all memory tasks. Taking into account that the average participant age in the present study comes in at 37.17 with a *SD* of 14.69, while the participants in Farmer's et al. (2017) study had a mean age of 18.89 and a *SD* of 0.99, this age gap could also account for the poorer performance in the present study. Still, based on the present results, it cannot be suggested that there is a significant correlation between working memory performance and the log transformed reading times in Garden Path sentences.

The present study further set out to test whether the presence of punctuation in the form of a grammatically placed comma would facilitate the processing of visually presented Garden Path structures and act similarly to prosodic breaks in spoken utterances. To measure this, the post-hoc pairwise comparisons in Table 11 were performed. Both interactions featuring the same level of ambiguity turned out to be non-significant, DNP:DP ($X^2 = 0.78$, $p = 0.38$), GPNP:GPP ($X^2 = 3.29$, $p = 0.14$). Therefore, the results of the present study do not speak in favour of a comma helping in disambiguation when reading time is considered the measurement. At least two possibilities could account for this result, unexpected under the *Implicit Prosody Hypothesis* (Bader, 1998; Fodor, 1998, 2002). The first possibility revolves around the stimuli design and the regions of analysis. As it stands, the design of the stimuli was not appropriate to add an additional region to the analysis to control for spillover effects which are likely to be found when administering a word-by-word self-paced reading task (Carreiras. & Clifton , 2004). Spillover effects are the phenomenon of the expected reading time effects only occurring after the critical region (i.e. the disambiguating Region 5) due to a processing delay resulting from the incoming difficulty caused by the need to resolve the ambiguity. As the critical region of interest in the present study falls exactly on the disambiguating region, and due to the stimuli design making it impossible to add a uniform spillover region, the present study is incapable to check for spillover effects, which is where the potential facilitation by the addition of a comma might be measurable.

A different possibility is that in a paradigm which combines both punctuated and non-punctuated stimuli, the punctuation markers potentially lead to a subconscious struggle within the parser, bringing additional attention to the stimuli and by consequence processing potentially takes longer.

6.3 Good Enough Parsing approach

The Good Enough parsing approach supposes that a parser does not fully specify the input they receive. In the case of Garden Path structures, this phenomenon results in the parser not resolving the presented ambiguity which can be observed by asking the readers a follow-up question probing the thematic role of the second DP (Blache, 2016; Christianson et al., 2001, 2006; Ferreira et al., 2001; Ferreira & Patson, 2007; Traxler, 2014). Examples of this can be seen in Table 3, one of them repeated below in (10).

(10) *Example Garden Path sentence and follow up question.*

Sentence: When the man knitted the socks fell to the floor.
Question: Did the man knit the socks?
correctAnswer: No.

Due to the ambiguity not being worked out, the parser is likely to give an incorrect answer, in the example above, the Good Enough parsing approach expects significantly more wrongful "Yes" replies compared to a disambiguous stimulus.

The descriptive statistics pertaining to the Good Enough parsing analysis (Table 7, Figure 5) already drew a clear distinction between the number of correct replies to Garden Path stimuli when compared to disambiguous ones. For the disambiguous stimuli, 913 out 936 (97.5%) questions were answered correctly, while only 23 were answered incorrectly (2.5%). This is in stark contrast with the Garden Path stimuli, in whose case 345 out of 936 (36.9%) were answered correctly, while 591 (63.1%) were wrongfully answered.

The logistic regression model in Table 13 investigated the impact of a number of predictors on the chance of receiving a correct reply to these questions. The two Garden Path conditions were the two most important predictors to receiving a correct answer. Among the two, the non-punctuated Garden Path condition (*OR = 0.01,* B = - 4.57, *SE* = 0.43) (7a.) came out slightly more impactful than the punctuated Garden Path condition (*OR* = 0.01, B = - 4.33, *SE* = 0.43) (7b.) when compared to the disambiguous condition without punctuation (7d.). According to the model then, the chance of getting a correct reply when comparing the punctuated Garden Path condition to the non-punctuated disambiguous condition is about 100 times lower (1/0.01).

The logistic regression model also considered the level of delay before the onset of the question as a predictor, as it has been hypothesised that a longer delay would allow more room for competition between the incorrect lingering first interpretation and the reanalysed structure (Kim & Christianson, 2017). Taking a look at "delay short" (500ms) (*OR* = 0.68, B = -0.38, *SE* = 0.16) as a predictor, it turns out that the short delay resulted in a decrease in correct replies when compared to the longer delay of 1500ms. This initial result runs counter to the prediction that a longer delay would lead to fewer correct responses. As a consequence, delay was also included as a fixed effect in the post-hoc analysis, including only the ambiguous conditions. The reason for this decision was that the prediction on delay and competition was specific to ambiguity resolution. If there is no ambiguity, no lingering interpretations are predicted (Kim & Christianson, 2017). The post-hoc model (Table 14) found the predictor *delay short* (*OR* = 0.67, B = -0.4, *SE* = 0.16) as improving the fit, with the number of correct replies decreasing in the presence of a short delay.

These data run counter to the assumption that a longer delay will lead to more competition between the possibilities and by extension lead to more wrong answers as a function of delay. In fact, the opposite pattern emerged, and given a long delay between the end of the sentence and the onset of the question improved the performance, participants performed better than in the presence of a short delay. A potential reason for this deviation from the predicted results could again stem from the stimuli not being designed with a spillover region in mind. A consequence of this design might be that the resolution of the ambiguity had not been processed, even in terms of shallow parsing, when the question was announced. When facing a short delay, it is then possible that the participants had yet to parse the ambiguity at all, therefore leading to a high number of wrong responses. When faced with a long delay, participants presumably had enough time to shallowly parse the ambiguity resulting in a higher likelihood for a correct response relative to the short delay.

The post-hoc analysis further found backwards digit span score to significantly improve the model fit, signifying that more questions were answered correctly the higher a participant scored on this memory task.

The lack of a significant improvement of the fit through the addition of the interaction between backwards digit span task and delay implies that the difference in delay does not stem from a difference in working memory capacity.

6.4 Punctuation

The present study set out to investigate whether the presence of a comma would aid to disambiguate Garden Path sentences in a visual paradigm, similarly to how prosodic boundaries have been shown to act for auditory input, as hypothesised under the *Implicit Prosody Hypothesis* (Bader, 1998; Fodor, 1998, 2002). If commas disambiguated the sentences, it was predicted that the resulting facilitation in processing could be seen either in a decreased reading time in Region 5, or an increased number of correct responses for the punctuated Garden Path condition over the non-punctuated version. In terms of reading time, adding punctuation to a null model did not provide a better fit, indicating that the presence of a comma had negligible impact on the reading time of Region 5. Descriptively, the discrepancy came out at about 30ms quicker reading times in favour of the punctuated condition (1370.91ms vs. 1340.66ms).

Returning to Table 14, the post-hoc logistic regression model summary for the number of questions answered correctly, it has to be noted that the fixed effect of *Condition* did not improve the model fit over a null model. Seeing that the only differences between the levels of *Condition* in the post-hoc analysis was the presence or absence of a comma, punctuation did not impact the performance when answering Good Enough parsing questions significantly. Descriptively in this study's sample, the non-punctuated Garden Path stimuli were answered correctly 161 out of 468 times (34.4%), while the punctuated Garden Path stimuli were answered correctly 184 out of 468 times (39.3%).

Taking the results of both analyses together, the present study did not find evidence of punctuation impacting the resolution of temporary ambiguities in either reading times or comprehension. As previously mentioned, the possibility exists that the reading time effect may have been observable if the study accommodated a spillover region.

6.5 Working memory

The influence of working memory on sentence processing, and more specifically ambiguity resolution has been a topic of debate (Brothers et al., 2021; Fedorenko et al., 2006; Freed et al., 2017; Traxler, 2012; Traxler et al., 2005; Van Dyke et al., 2014). In the present study, the impact of working memory was predicted to be visible in the reading times of Region 5, the disambiguating region for the Garden Path stimuli (7a, b.), with higher working memory capacity being predicted to result in a longer reading time of Region 5. When adding either of the two working memory scores as a predictor of reading time to the linear mixed effect model, summarised in Table 10, it did not result in a better fit of the model. It can therefore not be concluded that working memory capacity impacted reading times in Region 5 and, by extension, that the disambiguating region was not parsed more proficiently if an individual performed better on the working memory tests. This null effect of working memory on reading time in the disambiguating region could potentially stem from the participants in the present study having a poor working memory performance when compared to the participants in Farmer et al. (2017). As mentioned in the *Limitations* section, the poor results in the present study could stem from the online nature of the working memory task. As such, participants were unable to receive feedback, nor could their attentiveness be checked. Either of these two reasons could have led to an impaired performance. Further, mean participant age of the present study exceeding that in the study by Farmer et al. (2017) by more than double is another potential explanation for the weaker performance of this study's participants as working memory is commonly accepted to decline with growing age (Klencklen et al., 2017). A final reason for the weak performance of this study's participants on the working memory tasks could be fatigue. The total experiment had a median duration of 33 minutes, with the working memory tasks following the reading time experiment. It is possible that by the time participants got to the working memory tasks, they were already exhausted. While it was technically not prohibited for participants to take a break, *Prolific* imposes a maximum duration on experiments hosted on its site, which when exceeded, excludes the participants. Taken together, the poor working memory test performance has potentially led to weaker effect sizes than normal and impacted the final models.

The impact of working memory capacity was additionally predicted to be observable in the interaction between working memory scores and the level of delay between the end of the stimulus sentence and the onset of the comprehension question in the Garden Path conditions (7a, b.).

When attempting to fit either the working memory scores to the model post-hoc analysis for Good Enough parsing, maintaining only the ambiguous data, only the backwards digit span task improved the model fit. (Table 14). The interaction between both effects remained an insignificant improvement, suggesting that working memory does not interact significantly with delay when it comes to response accuracy to comprehension questions following Garden Path structures. The study finds that a short delay leads to more questions being answered incorrectly, implying that a longer delay did not lead to enhanced competition between two alternatives. The prediction made prior to the experiment was that a longer delay would lead to fewer correct answers due to the possible interpretations competing for a longer time (Ferreira & Patson, 2007). The present finding could come out of the stimuli paradigm not accounting for spillover effects. In short, spillover effects are the phenomenon of predicted effects only being detectable in the following region due to a processing delay. The stimuli for this experiment were not uniform enough to check for spillover effects as a number of them finished on the disambiguating region. If this were the case, the short delay would be equivalent to essentially no delay between the resolution of the ambiguity and the onset of the question, which could potentially explain the discrepancy between the prediction and the findings. A further option for the null effects found is the previously poor performance on the working memory tasks, potentially reducing the effect to a degree they are no longer detectable.

7. Conclusion

The present study aimed to find answers to four main questions. First, it aimed to replicate the elicitation of Garden Path effects, such as predicted under the Garden Path approach (Frazier & Fodor, 1978). To assess these effects, temporarily ambiguous sentences, in the form of Late Closure Garden Path sentences, were contrasted to completely disambiguous sentences and the reading time in the potentially disambiguating region was measured. The data was then fitted into linear mixed effect models (Table 10), to allow for control of random effects. In line with the prediction, the findings of the present study are in favour of the Garden Path approach, with both temporarily ambiguous conditions taking significantly longer to read than their unambiguous counterparts.

This finding was a preliminary to answer the second question the present study focussed on, namely whether there was evidence for lingering interpretations as they are predicted under the Good Enough parsing approach (Blache, 2016; Christianson et al., 2001, 2006; Ferreira et al., 2001; Ferreira & Patson, 2007; Traxler, 2014). For there to be a lingering interpretation, the study first had to establish evidence for the occurrence of reanalysis. To evaluate the presence of a lingering interpretation, the answers given to comprehension questions following either temporarily ambiguous or fully unambiguous sentences were contrasted. They were fitted into a logistic regression model, modelling whether a question was answered correctly as the outcome (Table 13). The results point towards a major discrepancy between both types of ambiguity, with both ambiguous conditions decreasing the likelihood to receive a correct reply significantly. The performance on the questions was also worse when there was a short delay before the onset of the question as opposed to a long delay. This could be indicative of either the ambiguity not having been resolved in the short time frame or reanalysis not having been completed.

The third issue the study set out to investigate was whether written punctuation serves as a disambiguation element in the same way prosodic boundaries do, as predicted under the *Implicit*

Prosody Hypothesis (Bader, 1998; Fodor, 1998, 2002).To evaluate the impact of punctuation, both reading times in the disambiguating region and accuracy on the comprehension questions were considered. Punctuation did not improve the fit of the linear mixed effect model (summary in Table 10) nor did condition make a significant change in the post-hoc binary regression (Table 14). From this absence of a better fit, it has to be concluded that punctuation did not play a significant effect in helping with the resolution of temporary ambiguities taking into account reading times in the disambiguating region and response accuracy following Garden Path sentences. This runs counter to the prediction made under the *Implicit Prosody Hypothesis*.

The final question the present study attempted to explore was the relation between working memory on ambiguity resolution. For this, participants' working memory capacity was assessed using a reading span as well as a backwards digit span task. The participant’s scores were then subsequently attempted to be included in both the linear mixed effect model for reading time in the disambiguating region (Table 10) and the post-hoc binary regression for Good Enough parsing (Table 14). Only in the regression analysis did backwards digit span score improve the model, increasing accuracy response as a function of backwards digit span score. Reading span score did not lead to an improved fit for either model.The present experiment can therefore not claim to support either the *Shared*- or *Dedicated Resources* accounts completely. The results relating to working memory should be taken with a grain of salt, as the participants scored lower than expected which may have obscured some effects.

The comprehension questions were also presented with two levels of delay, predicted to change response accuracy. A longer delay was predicted to lead to enhanced competition between the lingering interpretations and by extension a lower response accuracy (Ferreira & Patson, 2007). Delay did significantly improve the fit of the post-hoc regression model for Good Enough parsing (Table 14), however in the opposite direction than predicted. As such, the present study did not find support for the competition between lingering interpretations.

8. Further Research

The results of the present study present themselves for further research. Two primary caveats for the study were the relatively low working memory scores and the absence of a viable spillover region. As such one option for future studies would be to replicate the present study but simultaneously further normalising the stimuli to include a consistent spillover region and improving the validity of the working memory tasks. This could be achieved by either conducting them in person, therefore being able to clarify or intervene if something goes wrong or by providing the participants with concise feedback during the trials. It can be expected that either of these steps improves the scores on the working memory tasks and may make their effects easier to be captured.

References

Bader, M. (1998). Prosodic Influences on Reading Syntactically Ambiguous Sentences. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in Sentence Processing* (pp. 1–46). Springer Netherlands. https://doi.org/10.1007/978-94-015-9070-9_1

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Blache, P. (2016, November). Good-enough parsing, Whenever possible interpretation: A constraint-based model of sentence comprehension. *CSLP-2016*. https://hal.archives-ouvertes.fr/hal-01907632

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*(2), 211–252.

Brothers, T., Hoversten, L., & Traxler, M. (2021). Bilinguals on the garden-path: Individual differences in syntactic ambiguity resolution. *Bilingualism: Language and Cognition*, *24*, 1–16. https://doi.org/10.1017/S1366728920000711

Carreiras M., & Clifton C., J. (2004). *The On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond*. Psychology Press. https://login.ezproxy.leidenuniv.nl:2443/login?URL=https://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=111170&site=ehost-live

Cheng, Y., Rothman, J., & Cunnings, I. (2021). Parsing preferences and individual differences in nonnative sentence processing: Evidence from eye movements. *Applied Psycholinguistics*, *42*, 129–151. https://doi.org/10.1017/S014271642000065X

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology*, *69*(5), 817–828. https://doi.org/10.1080/17470218.2015.1134603

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic Roles Assigned along the Garden Path Linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006a). Younger and older adults'" good-enough" interpretations of garden-path sentences. *Discourse Processes*, *42*(2), 205–238.

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006b). Younger and Older Adults' 'Good-Enough' Interpretations of Garden-Path Sentences. *Discourse Processes*, *42*(2), 205–238. https://doi.org/10.1207/s15326950dp4202_6

Drury, J. E., Baum, S. R., Valeriote, H., & Steinhauer, K. (2016). Punctuation and Implicit Prosody in Silent Reading: An ERP Study Investigating English Garden-Path Sentences. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/article/10.3389/fpsyg.2016.01375

Farmer, T. A., Fine, A. B., Misyak, J. B., & Christiansen, M. H. (2017). Reading Span Task Performance, Linguistic Experience, and the Processing of Unexpected Syntactic Events. *Quarterly Journal of Experimental Psychology*, *70*(3), 413–433. https://doi.org/10.1080/17470218.2015.1131310

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources☆. *Journal of Memory and Language*, *54*(4), 541–553. https://doi.org/10.1016/j.jml.2005.12.006

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/s0010-0285(03)00005-7

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Ferreira, F., & Chantavarin, S. (2018). Integration and Prediction in Language Processing: A Synthesis of Old and New. *Current Directions in Psychological Science*, *27*, 096372141879449. https://doi.org/10.1177/0963721418794491

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of Garden-Path Sentences: Implications for Models of Sentence Processing and Reanalysis. *Journal of Psycholinguistic Research*, *30*(1), 3–20. https://doi.org/10.1023/A:1005290706460

Ferreira, F., & Patson, N. D. (2007). The 'Good Enough' Approach to Language Comprehension: The 'Good Enough' Approach. *Language and Linguistics Compass*, *1*(1–2), 71–83.

https://doi.org/10.1111/j.1749-818X.2007.00007.x

Fodor, J. D. (1998). Learning To Parse? *Journal of Psycholinguistic Research*, *27*(2), 285–319. https://doi.org/10.1023/A:1023258301588

Fodor, J. D. (2002). *Prosodic Disambiguation In Silent Reading*. *32*, 21.

Frazier, L., & Clifton, C. (1996). *Construal*. MIT Press.

Frazier, L., & Clifton, C. (1997). Construal: Overview, Motivation, and Some New Evidence. *Journal of Psycholinguistic Research*, *26*, 277–295. https://doi.org/10.1023/A:1025024524133

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325. https://doi.org/10.1016/0010-0277(78)90002-1

Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language*, *26*(5), 505–526. https://doi.org/10.1016/0749-596X(87)90137-9

Freed, E. M., Hamilton, S. T., & Long, D. L. (2017). Comprehension in proficient readers: The nature of individual variation. *Journal of Memory and Language*, *97*, 135–153. https://doi.org/10.1016/j.jml.2017.07.008

Galkina, E. (2015). Processing of Garden-Path Sentences Containing Silent and Filled Pauses in Stuttered Speech: Evidence From a Comprehensive Study. *Theses and Dissertations*. https://scholarcommons.sc.edu/etd/3139

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425–430. https://doi.org/10.1111/1467-9280.00475

Gráf, T. (2014). Up the Garden Path. *Philologica 3*, 85-98. https://karolinum.cz/data/clanek/288/085-098_Philologica_3_2013_2785-5.pdf

Hirotani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, *54*(3), 425–443. https://doi.org/10.1016/j.jml.2005.12.001

Kerkhofs, R., Vonk, W., Schriefers, H., & Chwilla, D. J. (2008). Sentence processing in the visual and auditory modality: Do comma and prosodic break have parallel functions? *Brain Research*, *1224*, 102–118. https://doi.org/10.1016/j.brainres.2008.05.034

Kim, J. H., & Christianson, K. (2013). Sentence Complexity and Working Memory Effects in Ambiguity Resolution. *Journal of Psycholinguistic Research*, *42*(5), 393–411. https://doi.org/10.1007/s10936-012-9224-4

Kim, J. H., & Christianson, K. (2017). Working memory effects on L1 and L2 processing of ambiguous relative clauses by Korean L2 learners of English. *Second Language Research*, *33*(3), 365–388. https://doi.org/10.1177/0267658315623322

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*(1), 15–47. https://doi.org/10.1016/0010-0277(72)90028-5

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*(2), 153–194. https://doi.org/10.1006/jmla.1998.2620

Klencklen, G., Banta Lavenex, P., Brandner, C., & Lavenex, P. (2017). Working memory decline in normal aging: Is it really worse in space than in color? *Learning and Motivation*, *57*, 48–60. https://doi.org/10.1016/j.lmot.2017.01.007

McRae, K., Spivey-Knowlton, M., & Tanenhaus, M. (1998). *Modeling the Influence of Thematic Fit (and Other Constraints) in On-line Sentence Comprehension*. https://doi.org/10.1006/JMLA.1997.2543

Omaki, A., Lau, E. F., Davidson White, I., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00384

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, *47*(4), 1343–1355. https://doi.org/10.3758/s13428-014-0543-2

Raiford, S. E., Coalson, D. L., Saklofske, D. H., & Weiss, L. G. (2010). CHAPTER 2—Practical Issues in WAIS-IV Administration and Scoring. In L. G. Weiss, D. H. Saklofske, D. L. Coalson, & S. E. Raiford (Eds.), *WAIS-IV Clinical Use and Interpretation* (pp. 25–59). Academic Press. https://doi.org/10.1016/B978-0-12-375035-8.10002-3

Scheifer, G. R. (2021). Ambiguous Sentence Processing: Hyper-active Gap Filling and the Garden

Path Model. An Online Self-paced Reading Study. [Bachelor's thesis, Leiden University].

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, *86*(1), 142–164. https://doi.org/10.1016/S0093-934X(02)00542-4

Steinhauer, K. (2014). Event-related Potentials (ERPs) in Second Language Research: A Brief Introduction to the Technique, a Selected Review, and an Invitation to Reconsider Critical Periods in L2. *Applied Linguistics*, *20144*, 393–417. https://doi.org/10.1093/applin/amu028

Steinhauer, K., Alter, K., & Friederici, A. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*, 191–196. https://doi.org/10.1038/5757

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, *30*(3), 267–295. https://doi.org/10.1023/A:1010443001646

Traxler, M. J. (2012). *Introduction to psycholinguistics: Understanding language science /*. Wiley,.

Traxler, M. J. (2014). Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences*, *18*(11), 605–611. https://doi.org/10.1016/j.tics.2014.08.001

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct Attachment Is Not a Form of Lexical Ambiguity Resolution. *Journal of Memory and Language*, *39*(4), 558–592. https://doi.org/10.1006/jmla.1998.2600

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, *53*(2), 204–224. https://doi.org/10.1016/j.jml.2005.02.010

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, *131*(3), 373–403. https://doi.org/10.1016/j.cognition.2014.01.007

van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2000). Chapter 24 - Unrestricted Race: A New Model of Syntactic Ambiguity Resolution. In A. Kennedy, R. Radach, D. Heller, & J.

Pynte (Eds.), *Reading as a Perceptual Process* (pp. 621–648). North-Holland. https://doi.org/10.1016/B978-008043642-5/50029-2

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2009). Predictive Mechanisms in Idiom Comprehension. *Journal of Cognitive Neuroscience*, *22*, 1682–1700. https://doi.org/10.1162/jocn.2009.21293

Vincenzi, M. de. (1991). *Syntactic Parsing Strategies in Italian: The Minimal Chain Principle*. Springer Netherlands. https://doi.org/10.1007/978-94-011-3184-1

White, M., Rajkumar, R., Ito, K., & Speer, S. R. (2014). Eye tracking for the online evaluation of prosody in speech synthesis. In A. Stent & S. Bangalore (Eds.), *Natural Language Generation in Interactive Systems* (pp. 281–301). Cambridge University Press. https://doi.org/10.1017/CBO9780511844492.012

Zhou, H., Rossi, S., & Chen, B. (2017). Effects of Working Memory Capacity and Tasks in Processing L2 Complex Sentence: Evidence from Chinese-English Bilinguals. *Frontiers in Psychology*, *8*, 595. https://doi.org/10.3389/fpsyg.2017.00595

Appendix A - Experimental Stimuli

Sentences	Questions	Condition	Source
1. When the man knitted the socks fell to the floor.	Did the man knit the socks? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
2. When the man knitted, the socks fell to the floor.	Did the man knit the socks? – No	Punctuated Garden Path	Scheifer (2021, 2022)
3. When the man knitted the socks, the keys fell to the floor.	Did the man knit the socks? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
4. When the man knitted the socks the keys fell to the floor.	Did the man knit the socks? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
5. When the woman ate the cheese slowly melted.	Did the woman eat the cheese? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
6. When the woman ate, the cheese slowly melted.	Did the woman eat the cheese? – No	Punctuated Garden Path	Scheifer (2021, 2022)
7. When the woman ate the cheese, the butter slowly melted.	Did the woman eat the cheese? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)

8. When the woman ate the cheese the butter slowly melted.	Did the woman eat the cheese? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
9. When the doctors discussed the problem suddenly worsened.	Did the doctors discuss the problem? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
10. When the doctors discussed, the problem suddenly worsened.	Did the doctors discuss the problem? – No	Punctuated Garden Path	Scheifer (2021, 2022)
11. When the doctors discussed the problem, the patient suddenly worsened.	Did the doctors discuss the problem? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
12. When the doctors discussed the problem the patient suddenly worsened.	Did the doctors discuss the problem? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
13. When the suspect killed the witness hastily ran.	Did the suspect kill the witness? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
14. When the suspect killed, the witness hastily ran.	Did the suspect kill the witness? – No	Punctuated Garden Path	Scheifer (2021, 2022)
15. When the suspect killed the witness, the officer hastily ran.	Did the suspect kill the witness? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)

16. When the suspect killed the witness the officer hastily ran.	Did the suspect kill the witness? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
17. When the coach chose the team mockingly laughed.	Did the coach choose the team? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
18. When the coach chose, the team mockingly laughed.	Did the coach choose the team? – No	Punctuated Garden Path	Scheifer (2021, 2022)
19. When the coach chose the team, the assistant mockingly laughed.	Did the coach choose the team? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
20. When the coach chose the team the assistant mockingly laughed.	Did the coach choose the team? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
21. When the army attacked the civilians fiercely resisted.	Did the army attack the civilians? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
22. When the army attacked the civilians fiercely resisted.	Did the army attack the civilians? – No	Punctuated Garden Path	Scheifer (2021, 2022)
23. When the army attacked the civilians, the	Did the army attack the civilians? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)

defenders fiercely resisted.			
24. When the army attacked the civilians the defenders fiercely resisted.	Did the army attack the civilians? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
25. While the firefighters helped the cat carefully followed.	Did the firefighters help the cat? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
26. While the firefighters helped, the cat carefully followed.	Did the firefighters help the cat? – No	Punctuated Garden Path	Scheifer (2021, 2022)
27. While the firefighters helped the cat, the neighbor carefully followed.	Did the firefighters help the cat? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
28. While the firefighters helped the cat the neighbor carefully followed.	Did the firefighters help the cat? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
29. When the patrol followed the suspects suspiciously hastened.	Did the patrol follow the suspects? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
30. When the patrol followed, the suspects suspiciously hastened.	Did the patrol follow the suspects? – No	Punctuated Garden Path	Scheifer (2021, 2022)

31. When the patrol followed the suspects the car suspiciously hastened.	Did the patrol follow the suspects? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
32. When the patrol followed the suspects, the car suspiciously hastened.	Did the patrol follow the suspects? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
33. While the mother bathed the baby happily played.	Did the mother bathe the baby? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
34. While the mother bathed, the baby happily played.	Did the mother bathe the baby? – No	Punctuated Garden Path	Scheifer (2021, 2022)
35. While the mother bathed the baby the dog happily played.	Did the mother bathe the baby? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
36. While the mother bathed the baby, the dog happily played.	Did the mother bathe the baby? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
37. When the detective heard the rumors quickly dispersed.	Did the detective hear the rumors? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
38. When the detective heard, the rumors quickly dispersed?	Did the detective hear the rumors? – No	Punctuated Garden Path	Scheifer (2021, 2022)

39. When the detective heard the rumors, the crowd quickly dispersed?	Did the detective hear the rumors? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
40. When the detective heard the rumors the crowd quickly dispersed?	Did the detective hear the rumors? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
41. When the youngsters left the party quickly ended.	Did the youngsters leave the party? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
42. When the youngsters left, the party quickly ended.	Did the youngsters leave the party? – No	Punctuated Garden Path	Scheifer (2021, 2022)
43. When the youngsters left the party, the fun quickly ended.	Did the youngsters leave the party? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
44. When the youngsters left the party the fun quickly ended.	Did the youngsters leave the party? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
45. When the travelers read the sign slowly changed.	Did the travelers read the sign? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
46. When the travelers read, the sign slowly changed.	Did the travelers read the sign? – No	Punctuated Garden Path	Scheifer (2021, 2022)

47. When the travelers read the sign, their attitude slowly changed.	Did the travelers read the sign? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
48. When the travelers read the sign their attitude slowly changed.	Did the travelers read the sign? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
49. When the student understood the teacher happily smiled.	Did the student understand the teacher? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
50. When the student understood, the teacher happily smiled.	Did the student understand the teacher? – No	Punctuated Garden Path	Scheifer (2021, 2022)
51. When the student understood the teacher the principal happily smiled.	Did the student understand the teacher? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
52. When the student understood the teacher, the principal happily smiled.	Did the student understand the teacher? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
53. While the spectators watched the match gradually escalated.	Did the spectators watch the match? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
54. While the spectators watched, the match gradually escalated.	Did the spectators watch the match? – No	Punctuated Garden Path	Scheifer (2021, 2022)

55. While the spectators watched the match the discussion gradually escalated.	Did the spectators watch the match? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
56. While the spectators watched the match, the discussion gradually escalated.	Did the spectators watch the match? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
57. While the teacher taught the class loudly roared.	Did the teacher teach the class? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
58. While the teacher taught, the class loudly roared.	Did the teacher teach the class? – No	Punctuated Garden Path	Scheifer (2021, 2022)
59. While the teacher taught the class, the animal loudly roared.	Did the teacher teach the class? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
60. While the teacher taught the class the animal loudly roared.	Did the teacher teach the class? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
61. While the daughter wrote the letter suddenly tore.	Did the daughter write the letter? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
62. While the daughter wrote, the letter suddenly tore.	Did the daughter write the letter? – No	Punctuated Garden Path	Scheifer (2021, 2022)

63. While the daughter wrote the letter, the paper suddenly tore.	Did the daughter write the letter? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
64. While the daughter wrote the letter the paper suddenly tore.	Did the daughter write the letter? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
65. When the mechanic pushed the car suddenly started.	Did the mechanic push the car? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
66. When the mechanic pushed, the car suddenly started.	Did the mechanic push the car? – No	Punctuated Garden Path	Scheifer (2021, 2022)
67. When the mechanic pushed the car, the engine suddenly started.	Did the mechanic push the car? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
68. When the mechanic pushed the car the engine suddenly started.	Did the mechanic push the car? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
69. When the man burned the wood loudly cracked.	Did the man burn the wood? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
70. When the man burned, the wood loudly cracked.	Did the man burn the wood? – No	Punctuated Garden Path	Scheifer (2021, 2022)

71. When the man burned the wood, the twigs loudly cracked.	Did the man burn the wood? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
72. When the man burned the wood the twigs loudly cracked.	Did the man burn the wood? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
73. While the woman hunted the deer quickly fled.	Did the woman hunt the deer? – No	No Punctuation Garden Path	Gráf (2014)
74. While the woman hunted, the deer quickly fled.	Did the woman hunt the deer? – No	Punctuated Garden Path	Gráf (2014)
75. While the woman hunted the deer, the rabbit quickly fled.	Did the woman hunt the deer? – Yes	Disambiguated Punctuated control	Gráf (2014)
76. While the woman hunted the deer the rabbit quickly fled.	Did the woman hunt the deer? – Yes	Disambiguated no punctuation control	Gráf (2014)
77. When the man baked the cake slowly cooled.	Did the man bake the cake? – No	No Punctuation Garden Path	Galkina (2015)
78. When the man baked, the cake slowly cooled.	Did the man bake the cake? – No	Punctuated Garden Path	Galkina (2015)

79. When the man baked the cake, the cookies slowly cooled.	Did the man bake the cake? – Yes	Disambiguated Punctuated control	Galkina (2015)
80. When the man baked the cake the cookies slowly cooled.	Did the man bake the cake? – Yes	Disambiguated no punctuation control	Galkina (2015)
81. While the bull charged the car inexplicably broke down.	Did the bull charge the fence? – No	No Punctuation Garden Path	Galkina (2015)
82. While the bull charged the car inexplicably broke down.	Did the bull charge the fence? – No	Punctuated Garden Path	Galkina (2015)
83. While the bull charged the fence, the tractor inexplicably broke down.	Did the bull charge the fence? – Yes	Disambiguated Punctuated control	Galkina (2015)
84. While the bull charged the fence the tractor inexplicably broke down.	Did the bull charge the fence? – Yes	Disambiguated no punctuation control	Galkina (2015)
85. While the accountant counted the money finally arrived.	Did the accountant count the money? – No	No Punctuation Garden Path	Galkina (2015)
86. While the accountant counted, the money finally arrived.	Did the accountant count the money? – No	Punctuated Garden Path	Galkina (2015)

87. While the accountant counted the money, the postman finally arrived.	Did the accountant count the money? – Yes	Disambiguated Punctuated control	Galkina (2015)
88. While the accountant counted the money the postman finally arrived.	Did the accountant count the money? – Yes	Disambiguated no punctuation control	Galkina (2015)
89. While the chauffeur drove the lady incessantly complained.	Did the chauffeur drive the lady? – No	No Punctuation Garden Path	Galkina (2015)
90. While the chauffeur drove the lady incessantly complained.	Did the chauffeur drive the lady? – No	Punctuated Garden Path	Galkina (2015)
91. While the chauffeur drove the lady, her husband incessantly complained.	Did the chauffeur drive the lady? – Yes	Disambiguated Punctuated control	Galkina (2015)
92. While the chauffeur drove the lady her husband incessantly complained.	Did the chauffeur drive the lady? – Yes	Disambiguated no punctuation control	Galkina (2015)
93. While the instructor graded the students slowly entered the room.	Did the instructor grade the students? – No	No Punctuation Garden Path	Galkina (2015)

94. While the instructor graded, the students slowly entered the room.	Did the instructor grade the students? – No	Punctuated Garden Path	Galkina (2015)
95. While the instructor graded the students, the parents slowly entered the room.	Did the instructor grade the students? – Yes	Disambiguated Punctuated control	Galkina (2015)
96. While the instructor graded the students the parents slowly entered the room.	Did the instructor grade the students? – Yes	Disambiguated no punctuation control	Galkina (2015)
97. When the carpenter measured the door spontaneously opened.	Did the carpenter measure the door? – No	No Punctuation Garden Path	Galkina (2015)
98. When the carpenter measured, the door spontaneously opened.	Did the carpenter measure the door? – No	Punctuated Garden Path	Galkina (2015)
99. When the carpenter measured the door, the window spontaneously opened.	Did the carpenter measure the door? – Yes	Disambiguated Punctuated control	Galkina (2015)
100. When the carpenter measured the door the window spontaneously opened.	Did the carpenter measure the door? – Yes	Disambiguated no punctuation control	Galkina (2015)

101. While the rider steered the pony profusely sweated.	Did the rider steer the pony? – No	No Punctuation Garden Path	Galkina (2015)
102. While the rider steered, the pony profusely sweated.	Did the rider steer the pony? – No	Punctuated Garden Path	Galkina (2015)
103. While the rider steered the pony, the spectators profusely sweated.	Did the rider steer the pony? – Yes	Disambiguated Punctuated control	Galkina (2015)
104. While the rider steered the pony the spectators profusely sweated.	Did the rider steer the pony? – Yes	Disambiguated no punctuation control	Galkina (2015)
105. When the customer ate the tomato fell on the ground.	Did the customer eat the tomato? – No	No Punctuation Garden Path	Galkina (2015)
106. When the customer ate, the tomato fell on the ground.	Did the customer eat the tomato? – No	Punctuated Garden Path	Galkina (2015)
107. When the customer ate the tomato, the fork fell on the ground.	Did the customer eat the tomato? – Yes	Disambiguated Punctuated control	Galkina (2015)
108. When the customer ate the tomato the fork fell on the ground.	Did the customer eat the tomato? – Yes	Disambiguated no punctuation control	Galkina (2015)

109. When the dog sniffed the owner came home.	Did the dog sniff the owner? – No	No Punctuation Garden Path	Galkina (2015)
110. When the dog sniffed, the owner came home.	Did the dog sniff the owner? – No	Punctuated Garden Path	Galkina (2015)
111. When the dog sniffed the owner, her friend came home.	Did the dog sniff the owner? – Yes	Disambiguated Punctuated control	Galkina (2015)
112. When the dog sniffed the owner her friend came home.	Did the dog sniff the owner? – Yes	Disambiguated no punctuation control	Galkina (2015)
113. While the therapist massaged the client loudly groaned.	Did the therapist massage the client? – No	No Punctuation Garden Path	Galkina (2015)
114. While the therapist massaged, the client loudly groaned.	Did the therapist massage the client? – No	Punctuated Garden Path	Galkina (2015)
115. While the therapist massaged the client, the assistant loudly groaned.	Did the therapist massage the client? – Yes	Disambiguated Punctuated control	Galkina (2015)
116. While the therapist massaged the client the assistant loudly groaned.	Did the therapist massage the client? – Yes	Disambiguated no punctuation control	Galkina (2015)

117. While the driver loaded the trunk just opened.	Did the driver load the trunk? – No	No Punctuation Garden Path	Galkina (2015)
118. While the driver loaded, the trunk just opened.	Did the driver load the trunk? – No	Punctuated Garden Path	Galkina (2015)
119. While the driver loaded the trunk, the suitcase just opened.	Did the driver load the trunk? – Yes	Disambiguated Punctuated control	Galkina (2015)
120. While the driver loaded the trunk the suitcase just opened.	Did the driver load the trunk? – Yes	Disambiguated no punctuation control	Galkina (2015)
121. While the artist painted the woman was impressed.	Did the therapist massage the client? – No	No Punctuation Garden Path	Galkina (2015)
122. While the artist painted the woman was impressed.	Did the therapist massage the client? – No	Punctuated Garden Path	Galkina (2015)
123. While the artist painted the woman, the curator was impressed.	Did the therapist massage the client? – Yes	Disambiguated Punctuated control	Galkina (2015)
124. While the artist painted the woman the curator was impressed.	Did the therapist massage the client? – Yes	Disambiguated no punctuation control	Galkina (2015)

125. While the maid attended the mistress became angry at the gardener.	Did the maid attend the mistress? – No	No Punctuation Garden Path	Galkina (2015)
126. While the maid attended, the mistress became angry at the gardener.	Did the maid attend the mistress? – No	Punctuated Garden Path	Galkina (2015)
127. While the maid attended the mistress, the housekeeper became angry at the gardener.	Did the maid attend the mistress? – Yes	Disambiguated Punctuated control	Galkina (2015)
128. While the maid attended the mistress the housekeeper became angry at the gardener.	Did the maid attend the mistress? – Yes	Disambiguated no punctuation control	Galkina (2015)
129. While the worker drilled the brick broke into pieces.	Did the worker drill the brick? – No	No Punctuation Garden Path	Galkina (2015)
130. While the worker drilled, the brick broke into pieces.	Did the worker drill the brick? – No	Punctuated Garden Path	Galkina (2015)
131. While the worker drilled the brick, the wall broke into pieces.	Did the worker drill the brick? – Yes	Disambiguated Punctuated control	Galkina (2015)

132. While the worker drilled the brick the wall broke into pieces.	Did the worker drill the brick? – Yes	Disambiguated no punctuation control	Galkina (2015)
133. While the lion chased the gazelle ran off.	Did the lion chase the gazelle? – No	No Punctuation Garden Path	Galkina (2015)
134. While the lion chased, the gazelle ran off.	Did the lion chase the gazelle? – No	Punctuated Garden Path	Galkina (2015)
135. While the lion chased the gazelle, the fawns ran off.	Did the lion chase the gazelle? – Yes	Disambiguated Punctuated control	Galkina (2015)
136. While the lion chased the gazelle the fawns ran off.	Did the lion chase the gazelle? – Yes	Disambiguated no punctuation control	Galkina (2015)
137. When the man dried the plate fell to the ground.	Did the man dry the plate? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
138. When the man dried, the plate fell to the ground.	Did the man dry the plate? – No	Punctuated Garden Path	Scheifer (2021, 2022)
139. When the man dried the plate, the spoon fell to the ground.	Did the man dry the plate? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)

140. When the man dried the plate the spoon fell to the ground.	Did the man dry the plate? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)
141. While the thief hid the jewelry dropped out of the bag.	Did the thief hide the jewelry? – No	No Punctuation Garden Path	Scheifer (2021, 2022)
142. While the thief hid, the jewelry dropped out of the bag.	Did the thief hide the jewelry? – No	Punctuated Garden Path	Scheifer (2021, 2022)
143. While the thief hid the jewelry, a ring dropped out of the bag.	Did the thief hide the jewelry? – Yes	Disambiguated Punctuated control	Scheifer (2021, 2022)
144. While the thief hid the jewelry a ring dropped out of the bag.	Did the thief hide the jewelry? – Yes	Disambiguated no punctuation control	Scheifer (2021, 2022)

Appendix B - Fillers

"No" Fillers	"No" Questions
1. When Bert came home, Greta greeted him.	Did Greta come home? – No
2. While Charles peeled the carrots, Melanie washed the potatoes.	Did Melanie wash peel the carrots? – No
3. When the woman performed her act, the spectators watched with attention.	Did the woman watch with attention? – No
4. While Mary drove the car, Ben talked without interruption.	Did Ben drive the car? – No
5. When Max passed the ball, Flavio missed the shot.	Did Flavio pass the ball? – No
6. When Berta threw the ball, the dog chased after it.	Did the dog throw the ball? – No
7. While the police investigated the crime, the suspect fled the country.	Did the police flee the country? – No
8. While the chef prepared the vegetables, the customers lost patience.	Did the chef lose his patience? – No
9. When the man fell down the stairs, the woman could not help but laugh.	Did the woman fall down the stairs? – No
10. While the children were playing, the parents watched them.	Did the parents play? – No
11. The seagull stole the food from the man's hand.	Did the seagull steal the man's drink? – No
12. The customer ordered pasta and wine.	Did the customer order a steak? – No

13. The runner started to doubt if she could reach the finish line.	Was the runner confident? – No
14. While the scientist conducted the experiment, the assistant recorded the data.	Did the scientist record the data himself? – No
15. When Sherlock arrived in Baker Street, Batman was already waiting for him.	Was Watson waiting in Baker Street? – No
16. The woman asked the man to marry her.	Did the man ask the woman to marry her? – No
17. When Gustav arrived at the party, the drinks were already out.	Were there any drinks left when Gustav arrived? – No
18. Why the baby screamed, the father did not know.	Did the father scream? – No
“Yes” Fillers	“Yes” Questions
1. When the song played on the radio, the family sang along.	Did the family sing along? – Yes
2. The driver took a turn to the left after the intersection.	Did the driver turn to the left? – Yes
3. The girl finished the Bachelor’s thesis to become a graduate.	Did the girl finish the Bachelor’s thesis? – Yes
4. The crowd could not fathom how the magician vanished.	Did the magician vanish? – Yes
5. The conductor did not know why the train stopped.	Did the train stop? – Yes
6. The landlord did not know that his package arrived.	Did the landlord’s package arrive? – Yes

7. The mother did the housework while thinking about her next vacation.	Did the mother do the housework? – Yes
8. The husband did the laundry while singing YMCA.	Did the husband sing YMCA? – Yes
9. The athlete did the exercise while sweating profusely.	Was the athlete sweating while doing the exercise? – Yes
10. When the glass hit the floor, it made a loud sound.	Did the glass hit the floor? – Yes
11. When the tram arrived the travelers were relieved they could finally hop in.	Did the tram arrive? – Yes
12. The rabbit jumped when it suddenly spotted the dog.	Did the rabbit spot the dog? – Yes
13. When Charlotte entered the ring, she became a different person.	Did Charlotte become a different person when entering the ring? – Yes
14. While the artist finished the painting, the brush fell apart.	Did the brush fall apart? – Yes
15. While the home burned down, the family escaped with their pets.	Did the family escape with their pets? – Yes
16. The astronaut waved at friends and family before entering the rocket.	Did the astronaut wave at friends and family? – Yes
17. When the sun rose, the adolescents were tired from sitting around the fire.	Were the adolescents tired? – Yes
18. When the star Basketball player missed a free throw, the crowd booed him.	Did the star player miss a free throw? – Yes

Appendix C - Background Questionnaire

How old are you?

What is your dominant hand?

- Right
- Left
- Ambidextrous

What is your country of origin?

Are you a native speaker of English?

- Yes
- No

What languages do you speak at home?

If you have acquired a second language, at what age did you start doing so? If not, please indicate "N.A."

Please rate your own proficiency in the individual languages you have acquired.

Do you suffer from any conditions which could affect your language comprehension? (i.e. dyslexia)

- Yes
- No

Appendix D - Consent Form

Hello, and welcome to the experiment! Before you start please read the participant information form.

Participant Information Form

Dear Participant, we would like you to take part in a study in which we hope to gain more understanding about sentence processing. We aim to do this by presenting sentences and asking comprehension questions about these sentences and relating the data to working memory tests.

This experiment is part of the MA thesis in Linguistics of the experimenter Gilles R. Scheifer, supervised by Dr. Leticia Pablos Robles at Leiden University.

Procedure of the experiment

The experiments consists of four parts, and will take a total of about 20 minutes. The first task is a self-paced reading task, where you manage the pace yourself by clicking the **Next** button to continue. After each sentence a comprehension question will follow. Following the reading task, two memory tasks will have to be done. Each task is preceded by instructions and trial sets. At the end of the memory tests, you will be shown your performance for the respective test.

Please focus on the tasks at hand, and try to minimise distractions.

Voluntary Participation

Participation in this study is voluntary and discretional. This means you can choose to withdraw at any time, without providing a reason. Do keep in mind though that you will not be paid if you decide to withdraw prematurely.

Participant Confidentiality

All information collected with regard to this study will be treated strictly confidentially. All data will be processed and stored anonymously. The data will not be available to unauthorised people and will not allow individual participants to be personally identified. This research is coordinated by Gilles R, Scheifer. Please contact him if you have any questions or comments about this study. You can find his contact information below.

Complaints

Should you find that you have been incorrectly or insufficiently informed about participation in this study, or if you have any complaints about the way this study was performed or the way you have been treated as a participant, it is recommended to discuss this with the experimenter or the coordinator of this study. If you do not wish to do so, or in case that does not resolve the issue, you can also lodge a complaint with the Leiden University Centre for Linguistics (LUCL). Please find LUCL's contact information below.

Consent

In order for you to participate in this study, we require your consent. You can confirm your willingness to participate by checking the box in the **Informed Consent** section below.

Contact Information

Experimenter: Gilles R. Scheifer

E-mail: g.r.scheifer@umail.leidenuniv.nl

Supervisor: Dr. Leticia Pablos Robles

E-Mail: l.pablos.robles@hum.leidenuniv.nl

Leiden University Centre for Linguistics (LUCL):

E-mail: lucl@hum.leidenuniv.nl

Informed Consent

By checking the box below, you confirm that you have read and understood the participant information form. You also confirm that you agree to the study procedure described in the information form.

I have read and understood the participant information form and I agree to participate in this study.

◯ I agree to participate in the study.
◯ I disagree to participate in the study.

Appendix E - Closing Words

Thank you for participating in this experiment!

I am interested in your feedback and thoughts, so please feel free to leave them in the text box below.

This is not a mandatory part of the experiment, but do make sure to click *Next* at the bottom of the page to wrap up the experiment and return to Prolific.

What do you think this experiment is about?

Do you have any feedback? (i.e. Where the instructions clear, was the font too small, etc.)

If you have additional questions or comments, feel free to e-mail me:

g.r.scheifer@umail.leidenuniv.nl

That is all, thank you for your time and have a nice day!

Next

Next

Appendix F - Mean reading time and SD per region per condition

Region	Mean(in ms)	SD	Condition
1	1017.36	410	GPP
1	1005.78	408.04	GPNP
1	1007.56	389.91	DP
1	996.03	411.02	DNP
2	542.92	236.83	GPP
2	529.73	210.09	GPNP
2	535.05	221.2	DP
2	533.52	224.12	DNP
3	1170.45	458.58	GPP
3	1003.9	440.11	GPNP
3	1145.33	485.93	DP
3	1096.75	443.51	DNP
4	N.A.	N.A.	GPP
4	N.A.	N.A.	GPNP
4	1134.8	418.2	DP
4	1150.78	434.01	DNP