Interference Based on Semantic Cues: Investigating Retrieval Interference Effects in Filler - Gap Dependencies

Fani Karageorgou

Doctor of Philosophy

University of York
Language and Linguistic Science
June 2024

Abstract

Interference phenomena in filler – gap dependencies have been attributed to similarity based retrieval interference (Lewis & Vasishth, 2005; Lewis et al., 2006; McElree, 2000, 2006; McElree et al., 2003; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2011). Two seminal studies provided findings claiming that those interference effects could be either due to semantically plausible distractor words that are sentence external (Van Dyke and McElree, 2006) or even sentence internal in which case they also lead to plausibility illusions for otherwise implausible sentences (Cunnings and Sturt, 2018). Two sets of experiments were run to understand those interference phenomena better. The findings of this thesis indicate that i. semantic similarity alone cannot always explain filler – gap dependency parsing, ii. structural information is also important and iii. semantic attraction is not just a repair mechanism for implausible sentences.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

Contents

1 **1**

1.1 Introduction 1
1.2 Long distance dependencies 1
1.2.1 Parsing long distance dependencies 3
1.2.2 Interference effects in sentence processing 9
1.3 Working memory 17
1.3.1 Models of working memory 17
1.3.2 Working memory and Language 22
1.4 Cue-based Retrieval 24
1.4.1 Activation-based model 24
1.4.2 Direct access model 30
1.5 Similarity-based Interference 31
1.5.1 Proactive vs. Retroactive Similarity-based Interference 32
1.5.2 Facilitatory vs. Inhibitory Similarity-based Interference 33
1.5.3 Interference in stages: encoding vs. retrieval 35
1.6 Retrieval interference effects in filler-gap dependencies 37
1.6.1 Proactive Inhibitory Retrieval Interference 38
1.6.2 Retroactive Facilitatory Retrieval Interference 42
1.6.3 Outline of Present Thesis 45

2 **46**

2.1 Similarity based retrieval interference from sentence external distractor words 46
2.1.1 Dependency resolution 46
2.1.2 Cue-based models 47
2.1.3 Sentence external retrieval interference 49
2.2 Pre-tests and Relatedness 51
2.3 Experiment 1: Cloze Task 52
2.3.1 Participants 52
2.3.2 Materials 52

2.3.3 Procedure . . . 53
2.3.4 Analysis . . . 53
2.3.5 Results . . . 54
2.3.6 Discussion . . . 54
2.4 Experiment 2: Lexical Decision Task . . . 55
2.4.1 Participants . . . 55
2.4.2 Materials . . . 55
2.4.3 Procedure . . . 55
2.4.4 Analysis . . . 56
2.4.5 Results . . . 56
2.4.6 Discussion . . . 57
2.5 Thematic Integration Account . . . 58
2.6 Experiment 3: Dual Task Self-paced Reading . . . 60
2.6.1 Participants . . . 60
2.6.2 Materials . . . 60
2.6.3 Procedure . . . 62
2.6.4 Analysis . . . 62
2.6.5 Regions of Interest . . . 64
2.6.6 Results . . . 64
2.6.7 Discussion . . . 66
2.7 Experiment 4: Maze Task I . . . 68
2.7.1 Participants . . . 68
2.7.2 Materials . . . 68
2.7.3 Procedure . . . 69
2.7.4 Analysis . . . 70
2.7.5 Regions of Interest . . . 70
2.7.6 Results . . . 70
2.7.7 Discussion . . . 72
2.8 Experiment 5: Simple Self-paced Reading Task . . . 74
2.8.1 Participants . . . 74
2.8.2 Materials . . . 74
2.8.3 Procedure . . . 75
2.8.4 Analysis . . . 75
2.8.5 Regions of Interest . . . 75
2.8.6 Results . . . 75
2.8.7 Discussion . . . 77
2.9 Effect of Task on Relatedness . . . 79
2.10 General Discussion . . . 80
2.10.1 Offline Effects . . . 80

2.10.2 Online Effects . . . 80
2.10.3 Theoretical Accounts . . . 82

3 **85**

3.1 Similarity based retrieval interference and semantic attraction from sentence internal distractor words . . . 85
3.1.1 Attraction effects . . . 85
3.1.2 Illusions of grammaticality . . . 87
3.1.3 Present set of studies . . . 88
3.2 Experiment 6: Plausibility Study for Maze Task II . . . 90
3.2.1 Participants . . . 90
3.2.2 Materials . . . 90
3.2.3 Procedure . . . 92
3.2.4 Analysis & Results . . . 92
3.3 Experiment 7: Maze Task II . . . 94
3.3.1 Participants . . . 94
3.3.2 Materials . . . 94
3.3.3 Procedure . . . 96
3.3.4 Analysis . . . 96
3.3.5 Regions of Interest . . . 97
3.3.6 Results . . . 97
3.3.7 Discussion . . . 99
3.4 Experiment 8: Maze Task III . . . 101
3.4.1 Participants . . . 101
3.4.2 Materials . . . 101
3.4.3 Procedure . . . 103
3.4.4 Analysis . . . 103
3.4.5 Regions of Interest . . . 103
3.4.6 Results . . . 104
3.4.7 Discussion . . . 105
3.5 Experiment 9: Rapid Serial Visual Presentation Task I . . . 107
3.5.1 Participants . . . 107
3.5.2 Materials . . . 107
3.5.3 Procedure . . . 108
3.5.4 Analysis . . . 109
3.5.5 Results . . . 109
3.5.6 Discussion . . . 111
3.6 Experiment 10a: Plausibility Study I for Maze Task IV . . . 113
3.6.1 Participants . . . 113

3.6.2 Materials . . . 114
3.6.3 Procedure . . . 114
3.6.4 Analysis & Results . . . 114
3.7 Experiment 10b: Plausibility Study II for Maze Task IV . . . 115
3.7.1 Participants . . . 115
3.7.2 Materials . . . 115
3.7.3 Procedure . . . 115
3.7.4 Analysis & Results . . . 116
3.7.5 Discussion . . . 116
3.8 Experiment 11: Maze Task IV . . . 117
3.8.1 Participants . . . 117
3.8.2 Materials . . . 117
3.8.3 Procedure . . . 117
3.8.4 Analysis . . . 118
3.8.5 Regions of Interest . . . 118
3.8.6 Results . . . 119
3.8.7 Discussion . . . 121
3.9 Experiment 12: Rapid Serial Visual Presentation Task II . . . 122
3.9.1 Participants . . . 122
3.9.2 Materials . . . 122
3.9.3 Procedure . . . 123
3.9.4 Analysis . . . 123
3.9.5 Results . . . 123
3.9.6 Discussion . . . 125
3.10 Effect of Experiment on Distractor Plausibility . . . 127
3.11 General Discussion . . . 129
3.11.1 Offline Effects . . . 129
3.11.2 Online Effects . . . 129
3.11.3 Theoretical Accounts . . . 130

4 **133**
4.1 Results and Cue-based Retrieval Models . . . 133
4.2 Effect of Task . . . 135
4.3 A Different Perspective . . . 136
4.4 Future Directions & Conclusion . . . 137

5 Appendices **138**
5.1 Appendix for tables . . . 138
5.2 Appendix for graphs . . . 143

5.3 Appendix for experimental stimuli 152
5.3.1 Experiment 3 stimuli 152
5.3.2 Experiment 4 stimuli 160
5.3.3 Experiment 5 stimuli 169
5.3.4 Experiment 7 stimuli 175
5.3.5 Experiment 8 stimuli 187
5.3.6 Experiment 9 stimuli 193
5.3.7 Experiment 11 stimuli 200
5.3.8 Experiment 12 stimuli 210

Chapter 1

1.1 Introduction

This chapter includes an overview of the theoretical and experimental concepts on long distance dependencies, how those are parsed via cue-based retrieval mechanisms and findings attributed to interference effects. The chapter is organised as follows: Section 1.2 presents examples of long-distance dependencies and discusses parsing strategies based both on syntactic as well as interference information. Having seen how parsing strategies require the use of working memory, section 1.3 then gives an overview of working memory models and discusses the close connection of working memory with language. Section 1.4 presents the mechanism of cue-based retrieval which incorporates memory concepts in linguistic processing accounts. Section 1.5 discusses a phenomenon that the cue-based retrieval accounts can very well explain, that is similarity based interference phenomena in different linguistic dependencies. Finally, section 1.6 presents the two studies that were the starting point of this thesis (Van Dyke and McElree (2006) and Cunnings and Sturt (2018)) and an outline of the rest of the thesis.

1.2 Long distance dependencies

A very important element in understanding a sentence is constructing the underlying syntactic structure of that sentence. Without this syntactic construction, successful comprehension would be rather difficult if not totally impossible (**i.e.**; Ferreira & Clifton Jr, 1986; Frazier & Fodor, 1978; Gibson, 1998; Stabler, 2015; Tanenhaus & Donnenwerth-Nolan, 1984). For instance without the construction of the different syntactic structure in the following two sentences a. and b. given in example 1, it would not be possible to interpret them. Constructing a parse tree for each of the sentences as the ones in figures 1.1 and 1.2, allows for their full interpretation and for understanding the different relationships (i.e., thematic roles, complements of the verb, etc.) between the different words.

(1) a. The woman saw the man.

b. The man saw the woman.

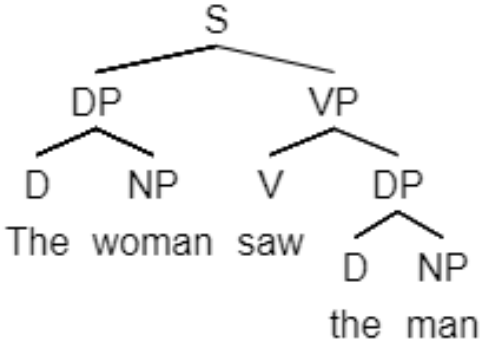


Figure 1.1: *Potential parse tree of sentence a. in figure 1.*

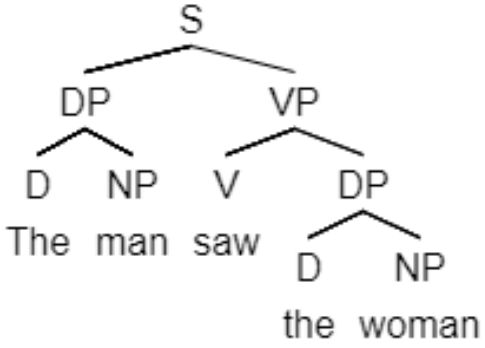


Figure 1.2: *Potential parse tree of sentence b. in figure 1.*

Multiple studies have shown that either when someone reads or hears a sentence, but in any case when one attempts to comprehend a sentence, parsing is largely incremental (**i.e.**; Altmann & Kamide, 1999; Frazier & Fodor, 1978; Garnsey et al., 1997; Just & Carpenter, 1980; Kamide et al., 2003; MacDonald et al., 1994; W. Marslen-Wilson, 1973; W. D. Marslen-Wilson, 1975; M. J. Pickering & Traxler, 1998; Tanenhaus et al., 1995; Trueswell et al., 1994; Tyler & Marslen-Wilson, 1977). That is people attempt to incorporate in a syntactic structure words as they read or hear them and do not wait until the end of that sentence before constructing the underlying syntactic structure of that given sentence. In the case of the sentences in 1, by incorporating the words to the parse tree linearly, as one reads or hears them, one might end up with two parse trees like the ones given in figures 1.1 and 1.2. Those parse trees are very revealing of the relationship between the different words and aid with the comprehension of those sentences, as they suggest that in the case of sentence a. in example 1 *the woman* could be the agent of the word *saw* (so the one doing the *"seeing"* in the sentence) and *the man* is the complement of that verb *saw*, whereas for the sentence b. in example 1 *the woman* is the complement of the verb *saw* and here *the man* is the one that is the agent. Therefore in those two cases incorporating the words linearly, in the way that they are read or heard, allows for the achievement of informative parsing and therefore successful comprehension of the sentences' meaning.

Not all sentences however could be processed by involving just incremental combination of different linguistic units. In many cases, parsing a sentence requires the formation of grammatical relationships between items (either words or even whole phrases) that are non-adjacent; and that is a case of a 'long-distance' dependency. Those are dependencies that require linking elements that are not presented linearly.

An example of a sentence involving a long-distance dependency is the sentence in example 2. In the example sentence in 2 for instance, the word *book* has to somehow be connected to the word *gave*. However those two elements are not linearly presented in the sentence but they are separated by other intervening material (i.e., *that my best friend's colleague*.

(2) a. The book that my best friend's colleague gave me last month with the blue letters on the cover was very nice.

In this case, an attempt to parse the sentence incrementally would appear rather futile as it would be impossible to restore the dependency between the two elements that are not linearly next to each other. The ability to construct that dependency, which is non-local, is a key feature of natural language, as the parser constantly comes across such structures.

1.2.1 Parsing long distance dependencies

There are many parsing models and strategies in the literature that have proposed ways in which the parser might process dependencies while one encounters a sentence, whether that is via reading or hearing (i.e., two stage accounts, frequency-based accounts, interactive accounts, syntax-first models, etc.). In this section though, the focus will be on a set of three theories that I deem to be the most relevant in the discussion of the phenomena investigated in this thesis.

Active Filler Strategy

One of the most researched and widely accepted parsing strategies for long distance dependencies that has been supported by a multitude of research findings in the literature is that of the active filler (Clifton Jr & Frazier, 1989; Crain & Fodor, 1985; Fodor, 1978; Frazier, 1987a, 1987b; Frazier & d'Arcais, 1989; Stowe, 1986). That is a strategy that is adopted when the parser encounters a very specific type of long distance dependencies, those that contain filler-gap constructions.

A filler-gap dependency is a dependency that sees two otherwise connected items being separated. Specifically, it involves an element that has been displaced from its original canonical position (that element is considered to be the filler) (Fodor, 1978; H. Wanner & Maratsos, 1978), leaving a gap in the sentence that needs to be filled in order to achieve successful comprehension H. Wanner and Maratsos (1978). In those dependencies then, the parser needs to identify the trace where filler originated from and establish an integration relationship (Chomsky, 1981).

(3) a. Which book do you think Mary liked more?

In example 3, the phrase *which book* is found at the beginning of the sentence, despite it being the grammatical object of the verb *liked*. In this case, the filler is considered to be the phrase *which book*, which has moved to the front of the sentence, while its original position was right after the verb *liked*. The movement of the filler phrase has left a gap behind right after the verb then, where the filler has to be realised in order for successful parsing to take place. An illustration of the movement of the filler in the example sentence 3, its trace and its original position (gap) is shown in figure 1.3.

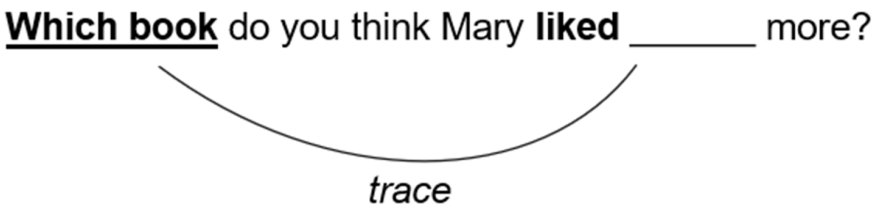


Figure 1.3: *An illustration of the filler-gap sentence in example 3.*

The active filler strategy posits that as soon as the parser encounters a filler, a displaced item, it starts actively seeking its corresponding gap so that it resolves that dependency. When that gap is found, the processor associates it with the filler, integrating the that filler with the corresponding verb and ultimately achieving an analysis of that sentence.

Notably, the active filler strategy seems to involve three distinct mechanisms during processing.

First, a predictive mechanism is assumed, as by the time the filler is encountered, the parser already predicts that a gap will be encountered later in the sentence as well (**i.e.**; Frazier & d'Arcais, 1989; Garnsey et al., 1989; Omaki et al., 2014; Stowe, 1986; Traxler & Pickering, 1996). Crain and Fodor, 1985 testing sentences with wh-movement argued that listeners predictively search for the gap location as soon as they encounter the wh-filler word. They also showed that participants experienced processing difficulties when the predictions they had made on the location of the gap were incorrect, indicating an active prediction mechanism. Frazier and Clifton Jr (1989) showed that readers often reanalyze sentences when their initial predictions about the location of an upcoming gap are proven wrong, with this reanalysis being taken as evidence for the active prediction of gap positions. Stowe (1986) conducting a series of experiments where the presence of gaps was manipulated found that readers showed increased reading times when gaps appeared in unexpected locations. Sentences like a. in example 4 had longer reading times at *us* compared to *us*

in sentence b. in example 4, as that could potentially be the first possible gap where the filler word *who* could be traced back to (if the word *who* was a complement of the verb *bring*). This finding supports the notion that readers actively predict gap locations while processing sentences.

(4) a. John wanted to know **who** Ruth will bring **us** home to ___ at Christmas.
    b. John wanted to know if Ruth will bring **us** home to mum at Christmas.

Traxler and Pickering (1996) in their research on filler-gap dependencies using eye -tracking, showed that readers posited the filler as soon as possible after encountering the sentence's verb. In a later study M. J. Pickering and Traxler (2003) testing sentences like the ones in the example below (example 5) in a self-paced reading design, showed that readers slow down when their predictions about gap locations are violated while at the same time processing is facilitated due to the expectation of a gap, supporting the active filler strategy. In the sentences in example 5, participants were faster at the word *worried* in sentence a. compared to the word *worried* in sentence b., as in the first sentence a gap could potentially be after *worried*, as the parser could assume that the *cat* is a complement of the word *worried*, while that is impossible for the word *car* in sentence b. (that is of course before one continues reading the sentence, where the word *about* signifies the position of the gap).

(5) a. That's the **cat** that the dog **worried** compulsively about after going to the vet.
    b. That's the **car** that the dog **worried** compulsively about after going to the vet.

Omaki et al. (2015) in a more recent study used visual world eye-tracking to demonstrate that participants predictively look at potential gap locations upon hearing a wh-filler, providing strong evidence for the predictive nature of the active filler strategy.

Apart from that, considering that filler-gap dependencies are long distance dependencies, the filler item has to be retained in memory until that point in the sentence when the gap is encountered (**i.e.**; Fujita & Cunnings, 2022; Ness & Meltzer-Asscher, 2017, 2019; Omaki et al., 2015; Wagers et al., 2012). That then requires that the filler is stored in working memory while the parser actively searches for its trace to its original position, while at the same time the parser processes other subsequent words (all the words in the sentence between the filler item and the gap). H. Wanner and Maratsos (1978) in their foundational study on sentence comprehension suggested that fillers are actively maintained in working memory until the gap is identified. They showed that participants experienced processing difficulties when the distance between the filler and the gap increased, implying a cognitive load associated with

maintaining the filler in memory. In a follow up study, E. Wanner et al., 1978 provided evidence that the processing difficulty increases with the distance between the filler and the gap as for example in the sentence in 6, suggesting that the filler is actively maintained in working memory. In the sentence in example 6, they found evidence that subjects attempted to posit the gap of the filler word *who* in multiple places (i.e., right after the words *ask*, *persuade*, *inform*, etc.), with each position being harder to process due to the maintenance of the filler word *who* in memory for longer until encountering the correct gap.

(6) a. Who did Tom ask Meg to persuade Jill to inform Ted that Bob had spoken to?

Phillips et al. (2005) also ran an ERP study testing wh-dependencies while manipulating their length. They found that the completion of a longer dependency correlates with a delayed P600. At the same time, they also found that an N400 component is elicited upon encountering verbs that the parser could initially assume to lead to the filler's gap, which however have a different argument structure leading to the non-integration of the filler word and therefore its continued maintenance.

Lastly, an integration mechanism is assumed as upon the encountering of the potential gap, the parser immediately attempts to integrate the filler it. In cases where that integration fails, for example in cases where the syntactic structure was incompatible or simply due to incorrect predictions, the parser triggers reanalysis processes as seen in many of the previous studies (**i.e**; Aoshima et al., 2004; Garnsey et al., 1989; Grodner & Gibson, 2005; Omaki et al., 2014; Phillips et al., 2005; M. Pickering & Barry, 1991; Stowe, 1986).

Minimal Chain Principle

A parsing strategy that could be similar to the active filler strategy discussed above just slightly modified when processing filler-gap dependencies, is the minimal chain principle (De Vincenzi, 1991, 2000).

According to this strategy, since the construction of syntactic chains is rather costly for short-term memory, when the parser detects that there is indeed a chain, it attempts to complete those chain computations as soon as possible.

Similar to the active filler strategy then, when the parser encounters a long-distance dependency, the minimal chain principle posits that the parser will attempt to establish the argument structure as soon as possible, so that items that have moved from their syntactic position are integrated back into their parse tree. It is suggested however that the minimal chain principle is not a parsing strategy just for long distance dependencies (De Vincenzi, 1991, 2000), but it could apply to the majority of

linguistic phenomena, by positing focus on the importance of not overloading the short term memory.

Based on this strategy then, a filler gap dependency like the example sentence in 3 would be resolved similarly to how the active filler strategy would predict, but for different reasons. The phrase *which book* is linked to the verb *liked* through a chain that is structured in the simplest and most direct manner possible, not necessarily looking for a trace and a gap but rather for the word that has a singleton chain that connects to the singleton chain of the phrase *which book*. The syntactic derivation of this sentence would involve movement operations that adhere to the minimal chain principle, ensuring that the chain linking *which book* to *liked* is minimal in terms of syntactic complexity.

Overall then, the minimal chain principle provides a theoretical foundation for understanding how syntactic dependencies are structured and by attempting to alleviate difficulties associated with more complex structures (i.e., long-distance dependencies) by following simplicity in structure. By emphasizing minimal chains, the principle contributes to understanding the importance of syntactic economy and its role in the prevention of short-term memory overload.

Dependency Locality Theory

Another parsing strategy specifically tailored to dependencies was proposed by Gibson and it is the dependency locality theory (Gibson, 1998; Gibson et al., 2000).

The theory taking into consideration that parsing sentences can be particularly costly, is shaped by two main elements that contribute to that cost. That structurally integrating items is costly for the parser (so connecting a word into the structure with the current input) while also at the same time keeping the syntactic structure of the input in memory (which might include keeping track of dependencies that are not complete). The theory then proposes that locality is the key that determines how dependencies are processed and the associated cost. For example it predicts that a longer distance between two elements (i.e., a pronominal referent and its antecedent, a filler item and its gap) will be associated with a higher cost and higher difficulty.

Based on this theory, a sentence like the one given in example 3 is expected to be processed in the way that would have the minimal cost. Therefore for that sentence the processor would be expected to posit the gap as early as possible, because that would significantly reduce the processing cost. It would mean that the filler (i.e., *which book* in the example sentence in 3) would not have to be maintained in memory any longer, awaiting for its gap and therefore its integration in a dependency. In that way the dependency locality theory appears to agree with the previously discussed theories in the way that filler-gap dependencies are processed.

The dependency locality theory can also account for other phenomena like the processing of object extracted vs. subject extracted relative clauses and much of research has focused on storage and integration costs that are associated with the processing of those structures (Gibson, 1998; Gibson & Ko, 1998; Grodner & Gibson, 2005; Hsiao & Gibson, 2003; Warren & Gibson, 2002).

The relative pronoun *who* in sentence a. in example 7 is extracted from the subject position of the relative clause, whereas the same pronoun (*who*) is extracted from the object position in sentence b. in the example in 7.

Gibson and Ko (1998) testing such sentences in a self-paced reading design offered support to the dependency locality theory, as it confirmed the theory's prediction. That is when processing the sentence b. in example 7 reading times should be higher for the word *sent* compared to the same word in sentence a. in example 7 as the cost for maintaining and then integrating the word *reporter* as the grammatical object is higher to the longer distance.

(7) a. **The reporter** who sent **the photographer** to the editor hoped for a good story.

b. **The reporter** who **the photographer** sent to the editor hoped for a good story.

Gibson and Ko (1998) suggested that object relative clauses like the one above in the example (7) are harder to process than subject relative clauses because both storage and integration costs are higher in object relatives. In sentence b. in the example in 7 the word *the reporter* not only has to be stored in memory for longer, but it also does while a new item is introduced later on in the sentence (that would be the word *the photographer*) before the word *reporter* can be integrated. That is not the case for a subject relative clause, as in sentence a. in example 7, the word *reporter* does not have to be maintained in memory for as long, while no new items are introduced in the parse.

The dependency locality theory also then suggests that storage and integration costs should be influenced by the number of new discourse referents that a syntactic dependency crosses. Warren and Gibson (2002) tested the claim in a series of complexity-rating experiments that used object relative clauses like the ones in the example 8 with varying types of referring expression between the relative pronoun and the relative clause verb, finding that the complexity of object relatives was lowest for indexical pronouns (that could be words such as the word *we*), which refer to discourse entities that are highly accessible Ariel, 1990. In contrast, complexity was high for indefinite and definite noun phrases, which tend to refer to discourse entities that are relatively inaccessible. Therefore it was concluded that sentence complexity is higher in cases where a dependency crosses a referring expression while

more processing effort is also required because it refers to a less accessible discourse entity.

(8) a. The consultant who **we / the chairman / a chairman / John Smith** called advised wealthy companies about tax laws.

Overall, all those three parsing strategies attempted to explain results of studies focused on different types of long distance dependencies. Most importantly though, all three showed that different pieces of information such as syntactic information, memory associated costs and locality are important when processing linguistic dependencies.

1.2.2 Interference effects in sentence processing

Considering the limitations of the previous accounts in how long-distance dependencies are processed, other researchers attempted to attribute the observed effects to interference. The limitation of the Active Filler Strategy is primarily that it relies very heavily on (generally structural) information about the displaced filler element and the location of the gap, and not putting a lot of emphasis on other sources of information (i.e., other syntactic, semantic, pragmatic) that could potentially affect the processing of the dependency. Similarly, the Minimal Chain Principle is focused purely on the processing cost, once again not taking into account other pieces of information. Ultimately, the Dependency Locality Theory cannot fully explain certain effects (i.e., findings in Warren and Gibson (2002)), exactly due to the fact that semantic / pragmatic or anything other than location is not a contributing factor to the easiness of parsing dependencies.

Multiple studies have looked into those effects in the comprehension of multiple kinds of dependencies suggesting that interference is a phenomenon that occurs in the presence of elements either before (proactively) or after (retroactively) a dependency which leads to either facilitation or inhibition of the dependencies' processing (Cunnings & Sturt, 2018; Dillon et al., 2013; Gordon et al., 2001, 2002, 2006; Kush & Phillips, 2014; Wagers et al., 2009).

Interference in object- vs. subject-extracted structures

Investigating interference effects in a very particular type of dependency, Gordon et al., 2001 tested object-extracted vs. subject-extracted embedded sentences in a self-paced reading design (9).

(9) a. The banker that **the barber / you / Joe** praised climbed the mountain.
b. The banker that praised **the barber / you / Joe** climbed the mountain.

They found that overall, despite participants initially replicating the commonly observed pattern of processing simple subject-extracted sentences more easily than their object-extracted counterpart ones, the observed difficulty in parsing would change depending on the distractor item in the second noun phrase involved in the sentences. When the sentence involved two noun phrases of the same type, in this case a description noun phrase (i.e., *the banker* as the first noun phrase and *the barber* as the second one), then participants were slower in reading and less accurate in their responses in the object-extracted sentences compared to the subject-extracted ones. However, when the sentence involved two noun phrases of a different kind (i.e., *the banker* as the first noun phrase and *you* as the second one), then that had an effect in processing, such that participants performed similarly in the task, as well as in their responses to comprehension questions in both object-extracted and subject-extracted sentences. The same effect (if not more robust) was also observed when testing cleft sentences which involved two proper names as opposed to two description noun phrases (10), with similar reading times and comparable levels of accuracy for both the object-extracted and the subject-extracted sentences, when those involved either two proper names or two description nouns.

(10) a. It was **the barber / John** that saw **the lawyer / Bill** in the parking lot.
b. It was **the barber / John** that **the lawyer / Bill** saw in the parking lot.

This elimination of the advantage during the processing of the subject-extracted sentences over the object-extracted ones was taken to be an effect of similarity based interference. The researchers suggested that the reason for the observed effects is simply the greater similarity between the cues of each of the same type items that need to be stored in memory compared to the features of elements that do no belong in the same class (i.e., description noun phrase - description noun phrase vs, description noun phrase - pronoun). The features of elements that belong in the same class tend to be more alike, which in turn complicates the parser more when trying to process object-extracted sentences like the ones in the study; an idea fully predicted and definitely expected in a cue-based framework world.

Gordon et al. (2004) extended this work, by testing similarly structured object-extracted and subject-extracted relative clauses as seen in the example sentences in 11, with the main difference being the inclusion of *everyone* as a potential second noun phrase.

(11) a. The salesman that **the accountant / everyone** contacted spoke very quickly.
b. The salesman that contacted **the accountant / everyone** spoke very quickly.

Including the quantified pronoun *everyone* also seems to eliminate the expected difference in the parsing of object-extracted vs. subject-extracted relative clauses, as

participants in the study had similar reading times (in a self-paced reading task) as well as identical levels of accuracy in comprehension questions following the critical sentences with *everyone*, whereas responses and reading times of sentences containing two proper noun noun phrases were significantly different. The elimination of the expected effect is attributed by Gordon et al. (2004) to similarity based interference, as the pronoun *everyone* by default cannot share similar referential features with proper nouns. Due to the lack of similar semantic cues between *everyone* and any proper noun, no interference effects can be observed in the sentences containing it, whereas similar semantic features are more common between two proper nouns.

In a subsequent study, Gordon et al. (2002) extended this previous work by testing object-extracted and subject-extracted cleft sentences in an interesting design. The sentences were tested in a dual task which involved memorising three words of the same class (either three description nouns which were all human roles in this study, or three proper names), followed then by a sentence which included two noun phrases that either matched or mismatched the three memorised items as seen in the example in 12.

(12) a. **Joel - Greg - Andy**
   a. It was Tony that **liked Joey / Joey liked** before the argument began.
   b. It was the dancer that **liked the fireman / the fireman liked** before the argument began.

   b. **poet - cartoonist - voter**
   a. It was Tony that **liked Joey / Joey liked** before the argument began.
   b. It was the dancer that **liked the fireman / the fireman liked** before the argument began.

When there was a match between the three memorised words and the two noun phrases in the sentence, accuracy to comprehension questions was impaired and on-line reading times of the region containing the second noun phrase were significantly longer. For instance in example 12b.b., they found longer reading times for the *fireman* compared to 12b.a. at the word *Joey*. That was taken to suggest that comprehension is impaired due to cue-based interference, as there is an overload of similar features which are shared among the five different items (three items in memory load and two in the sentence) when those happen to belong in the same class, and not due to the number of items that need to be processed.

Based on this work, Fedorenko et al. (2006) used a similar paradigm which involved a dual task of memorising one or three nouns (either occupation nouns or proper names) prior to reading object-extracted vs. subject-extracted relative clauses in a self-paced reading experiment. Example sentences are presented in example 13.

(13) a. The physician who **consulted the cardiologist** checked the files in his office.

b. The physician who **the cardiologist consulted** checked the files in his office.

Once again, the memorised nouns were either a match or a mismatch to the two noun phrases in the sentences. A significant online effect was reported for the object-clefted relative clauses, such that when they were preceded by three words that matched the type of the noun phrases in the sentence, then there were longer reading times for the second noun phrase; however, the offline result of Gordon et al. (2002) was reportedly not replicated here. Fedorenko et al. (2006) took that to suggest that when processing structurally more complex sentences, the presence of multiple items that potentially match the type of an element in a dependency, and therefore there is a higher probability of their features matching when storing them as vectors of cues, could lead to cue-based interference, which is manifested in the longer reading times for the region.

Interference in subject-verb agreement structures – Agreement attraction

Similar phenomena have long been tested across multiple languages, initially in production and then comprehension. A different type of dependency tested for interference effects, involves interference effects in subject-verb agreement violations, where intervening but structurally inaccessible distractor items erroneously agree with a verb in the sentence leading to so called agreement attraction effects that give the illusion of a plausible sentence.

Bock and Miller (1991) sought to investigate number agreement errors, where verbs wrongly agree in number with an item that is not the licensed subject. In a series of production experiments, participants were asked to produce endings for preambles which involved the manipulation of the number of the subject noun as well as the number of an intervening distractor item that was part of a prepositional phrase as seen in example 14.

(14) a. The **key** to the **cabinet / cabinets** is/are..

b. The **keys** to the **cabinet / cabinets** is/are..

The critical finding was that agreement errors between subject and verb tend to be induced by plural intervening distractor words when those mismatched in number with the subject noun. Plural distractors following singular subjects nouns tended to interfere with the plurality of the produced verb, suggesting that people elicit verbs that agree with another element and not with the actual subject in the sentence. In this case, participants appeared to match the feature of number from a local noun (the

distractor) with a produced verb despite that violating the basic structural agreement between subject and verb.

The same type of agreement attraction effects has also been investigated cross-linguistically, with other studies reporting similar results with the ones in Bock and Miller (1991). For instance in Italian, Vigliocco et al. (1995) in a sentence completion task tested preambles that involved a subject head noun followed by a second noun that was part of a prepositional phrase and manipulated the number of both while looking into the number of agreement errors in the elicited verbs. An example of the items used are in example 15.

(15) a. The **discovery** of the **scientist / scientists** __.
b. The **discoveries** of the **scientist / scientists** __.

In similar designs, Vigliocco et al. (1996) extended this work in Spanish (using materials like the ones in example 16) and Vigliocco (1996) found the same effects in Dutch and French. In all these studies, there were more attraction errors when the subject noun was singular and the intervening distractor noun was plural with participants reliably producing more plural verbs than when both the subject noun and the distractor noun were singular.

(16) a. The **road** to the **lake / lakes** __.
b. The **roads** to the **lake / lakes** __.

However, in the case of a plural subject noun such effects are not induced, with the probability of error being low and approximately equal independently from the number of the intervening noun. These findings are consistent and robust in many languages, with studies in Dutch (Bock et al., 2001; Hartsuiker et al., 1999, 2003; Kaan, 2002; Meyer & Bock, 1999), German (Hartsuiker et al., 2003; Häussler, 2009), Spanish (Franck et al., 2008; Vigliocco, 1996), French (Franck et al., 2002, 2008, 2010; Vigliocco & Franck, 2001; Vigliocco et al., 1995), Italian (Franck et al., 2006, 2008; Vigliocco & Franck, 2001; Vigliocco et al., 1995), Russian (Lorimor et al., 2008) and Slovak (Badecker & Kuminiak, 2007) all showing the same effects in relevant experiments.

Franck et al. (2002) further extended this finding of agreement attraction in production, by testing whether syntactic prominence, so if an element that is higher up in the tree structure, contributes to these phenomena as much as linear distance between an equivalent distractor and the 'attracted' verb, which is what was usually tested in the studies previously discussed. A production task run in both English and French involved preambles that contained three different nouns (a head noun, an intermediate noun and a local noun) with the manipulation of the number of each one of them, An example is given in (17).

(17) a. The **inscription** on the **door / doors** of the **toilet / toilets** __.

b. The **inscriptions** on the **door / doors** of the **toilet / toilets** __.

For both languages more agreement errors were reported in the mismatch between the intervening noun and the other nouns, suggesting that syntactic distance has a larger impact on agreement attraction than linear distance, as that appeared to induce more of these errors. That is as the intervening noun is higher in the hierarchical structure and therefore impacts the verb more than a noun that might be local, but is lower in the structure. Another important effect reported is the induced effects with plural head subject nouns, suggesting that plural head nouns also lead to interference effects as the induce more errors overall, an effect which was not observed in other studies.

Following the seminal study of Bock and Miller (1991), Pearlmutter et al. (1999) tested similar effects of agreement attraction based on number features. The main difference between this and the the previously discussed studies was the methods of testing and the subsequent processes involved, with Pearlmutter et al. (1999) looking into agreement effects in comprehension compared to the previous studies which investigated these effects in production. Sentences like the ones in example (18) were tested in a self-paced reading paradigm as well as in eye-tracking.

(18) a. The *key / keys* to the *cabinet / cabinets* **was** rusty from many years of disuse.

b. The *key / keys* to the *cabinet / cabinets* **were** rusty from many years of disuse.

Both experiments showed that there is facilitation in processing when the main verb does not agree with the grammatical subject but the second noun phrase in number. The self-paced reading study results showed a clear facilitation at the word following the critical verb (i.e., *rusty*), suggesting that despite the sentence not being fully grammatical, the second noun phrase in the sentence interferes with the dependency as it involves some cues matching the ones at the verb, and therefore facilitates parsing despite the violation of agreement. This result was confirmed with eye-tracking, which also showed a reliable effect for the critical verb as well. In the ungrammatical sentences that contained a main verb agreeing in number not with the subject noun but with the intervening noun which is part of the prepositional phrase, so in sentences like *The key to the cabinets were rusty from many years of disuse.* (example 18b), where the main verb *were* agrees in number with the intervening distractor word *cabinets* but not with the grammatical subject *key*, they found short fixations for *were* and for the word immediately following it (the word *rusty* in this example). The fixations in those two words in the ungrammatical sentences were shorter than the fixations for the same words (namely *were* and *rusty*) in sentences where there was no agreement with either of the two noun phrases in the sentence, so in sentences like this: *The key to the cabinet were rusty from many years of disuse.*(example 18b).

The results of both methods showed that during processing there is facilitation in the reaction times and in the duration of the total fixations in ungrammatical sentences (so in sentences where the verb does not agree in number with the subject noun), when there is an intervening distractor noun that matches in number with the verb (i.e., *key - cabinets - were*), a phenomenon that is in full agreement with the cue-based framework's predictions. At the same time, it is reported that this mismatch in agreement has no impact on the comprehension of these sentences as accuracy rates in the comprehension questions following these sentences were almost identical.

Wagers et al. (2009) also tested number agreement attraction effects in a series of self-paced reading paradigms with a particular focus again on these effects in comprehension rather than in production, similarly to Pearlmutter et al. (1999). Adapted sentences from Pearlmutter et al. (1999) like the ones given in example 19, replicated previously found results, in that intervening distractors can have a facilitatory effect on the processing of a matching verb even if that verb does not agree in number with its grammatical subject.

(19) a. The key to the *cell / cells* unsurprisingly **was** rusty from many years of disuse.

b. The key to the *cell / cells* unsurprisingly **were** rusty from many years of disuse.

Importantly, by making use of retroactive and proactive interference of both distractors in prepositional phrases, as well as in relative clause constructions, Wagers et al. (2009) identified agreement effects as a characteristic of ungrammatical sentences only, whereas they suggested that intervening distractors in grammatical sentences do not lead to such an attraction. Those results then suggest that based on a cue-based framework, there is a mechanism that in an attempt to parse otherwise unsalvageable sentences creates an illusion of grammaticality by utilising and retrieving features of elements that are not grammatically or syntactically correct but that however are a good match to the retrieval point in a dependency.

Laurinavichyute and von der Malsburg (2022) investigated attraction effects similar to the ones discussed in Pearlmutter et al. (1999) and Wagers et al. (2009) but rather than focusing on the grammatical agreement (or actually lack of) between two items that form a dependency and instead the agreement between an item that is part of a dependency and an intervening distractor item that attracts it by matching the number of the verb, they looked into attraction effects through the prism of semantic similarity. More specifically, they were interested in whether processing would be facilitated when intervening distractors items that could be a good semantic match to otherwise sentences that contained semantic violations between the items that formed the subject and verb dependency, in a similar fashion to the facilitation

of processing as a result of agreement attraction (as showcased in i.e., (Bock & Miller, 1991; Wagers et al., 2009), etc.). They tested sentences like those in example 20 in a forced choice paradigm, where participants were asked to judge whether the verb of the sentences (so the word *sprays* in example 20) was a plausible continuation of the sentence presented to them (i.e., *The bakery near the office building / fire hydrant rarely* in example 20). They found results consistent with attraction, such that participants were reliably less accurate when making judgements for the sentences containing a distractor item that semantically matched the sentences' verb. In essence, they found that when encountering *sprays* in example 20b, participants were significantly less accurate judging it as not a plausible continuation for the sentence compared to the judgement they make for *sprays* in example 20a. That was taken to be consistent with attraction effects, as it is explained by Laurinavichyute and von der Malsburg, 2022 to be a result of people making less accurate judgements due to the existence of a semantically related item in the sentence (i.e., *fire hydrant* being semantically related with *sprays* in 20b, compared to *office building* which is not semantically related with *sprays* in 20a). Having encountered the *fire hydrant* in 20b was thought to be the reason that participants thought that *sprays* is a plausible continuation and eventually led to lower accuracy (due to *sprays* not actually being a plausible continuation of either of the sentences in 20).

(20) a. The bakery near the **office building** rarely sprays.
b. The bakery near the **fire hydrant** rarely sprays.

That was taken to suggest that semantic attraction phenomena are very similar to agreement attraction ones, in that they are both processed by the parser as interference induced effects.

What all the previously presented models appear to share, is the idea that processing dependencies very heavily relies on memory processes, and even more specifically on working memory processes. All the previous accounts and theories (Active Filler Strategy, Minimal Chain Principle, Dependency Locality Theory, and Interference) assume that for the parser to process a long distance dependency it is essential to store some pieces of information in memory (whatever those might be depending on the parsing theory). And importantly not simply store that information but being in a position to retrieve it as well in order to process the dependency by integrating the displaced item to its original position. To better understand how memory works then, I will now give a brief historical overview of working memory and its connection to language.

1.3 Working memory

Language is one of the most important cognitive skills that enables humans to interact while conveying meaning in those interactions. Therefore, the ability to process linguistic input and output is essential for effective communication. Being able to process linguistic material successfully requires making use of one of the brain's most central and frequently used mechanisms, language. The role of memory during linguistic processing appears to be vital as discussed previously (Aoshima et al., 2004; Clifton Jr & Frazier, 1989; Crain & Fodor, 1985; Cunnings & Sturt, 2018; De Vincenzi, 1991, 2000; Dillon et al., 2013; Fodor, 1978; Frazier, 1987a, 1987b; Frazier & d'Arcais, 1989; Garnsey et al., 1989; Gibson, 1998; Gibson et al., 2000; Gordon et al., 2001, 2002, 2006; Grodner & Gibson, 2005; Kush & Phillips, 2014; Omaki et al., 2014; Phillips et al., 2005; M. Pickering & Barry, 1991; Stowe, 1986; Wagers et al., 2009), as certain operations have to occur so that the linguistic input gets processed successfully. Memory capacity and skills have to be utilised for processes that are relatively basic, such as knowing and using words to describe a certain thing, to more complex, such as using or understanding structures that are the bases of sentences. Therefore, memory appears to play a central role when dealing with language, both for recalling the names for different things as well as sentence processing.

1.3.1 Models of working memory

Memory is a cognitive system which has long been studied for its essential role in the way that humans function in their everyday interactions. A clear distinction between the subdivisions of this system is based on the separate and rather different purposes that they serve as well as their manifestation on the electrical activity in the brain Hebb (1949). The occupation of space, the duration of storing information and therefore the difference in the impact that the corresponding electrical brain activation of the neurons appear to be what draws the line between the different types of memory. Based on Hebb (1949), when a certain type of information is the cause of electrical activity that leads to long lasting changes in neurological structures, then it is stored or is part of the so called long-term memory. Whereas information causing temporary neuronal activity is part of the short-term or working memory.

In line with this early observation, R. C. Atkinson and Shiffrin (1968) proposed a model which attempted to explain and describe memory processes by defining two separate components that are involved: short- and long-term memory. An illustration of the model that was proposed is given in figure 1.4. As it becomes clear from the illustration, it was suggested that information deriving from stimuli in the environment is initially processed by a short-term or working memory component which

allows for it to be stored for only a short period of time until it is processed, while then this information either decays and subsequently frees the occupied working memory space, or it is then stored in long-term memory more permanently, especially if the information has been repeatedly attended or in the case it has been occupying space in working memory for a prolonged period of time. Support to such a dual component model had been provided by other studies (i.e., J. Brown, 1958) which showed experimentally that the decay of items stored in memory is very different depending on whether they are stored in short- or long-term memory, as items stored in short-term memory tend to decay, or be forgotten very quickly, especially when they are not repeatedly rehearsed.

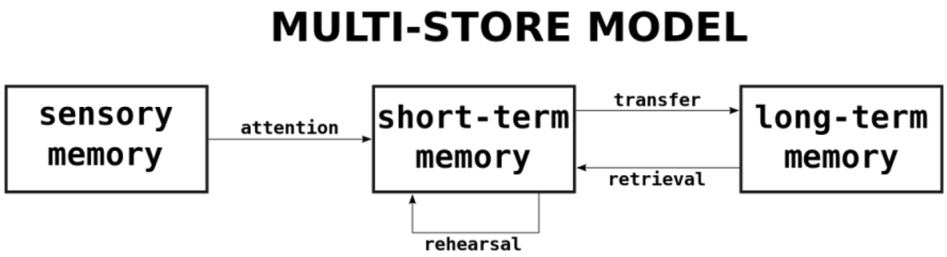


Figure 1.4: *An illustration of the memory model proposed by R. C. Atkinson and Shiffrin (1968).*

Since then, multiple studies have showed the central role that memory has in the processing of linguistic input (**i.e.**; A. Baddeley, 2003; Caplan & Waters, 2013; Caplan & Waters, 1999; Daneman & Merikle, 1996; Just & Carpenter, 1992; Miller & Isard, 1963). Successful comprehension requires access not only to long-term memory to retrieve stored items, but also to representations that have been briefly encountered earlier and are outside of the current focus of attention. That allows for inferences to be drawn, structural relations to be resolved and eventually sentences to make sense. Those representations that are not stored in long-term memory and at the same time are not in focus, are what constitutes stored items in working memory and this is where the focus of the present work will be placed from here onward.

In an attempt to understand the structure and the function of working memory better, A. D. Baddeley and Graham (1974) proposed a model rather different from what was suggested by R. C. Atkinson and Shiffrin (1968). In the R. C. Atkinson and Shiffrin (1968) model, short-term or working memory was presented in a simplistic view where it was treated as a unit; as a construct where a limited number of items can be held for a limited amount of time. In that model there was no suggestion however that working memory and short-term memory might refer to different processes, or that they could have multiple compartments responsible for different operations, but rather there was the idea that short-term and working memory are unified in their functions and processes and at the same time that the physical struc-

tures used were very similar to the ones used by long-term memory. A. D. Baddeley and Graham (1974) suggested that short-term memory and working memory are not the same and that there is a different structure from long-term memory.

Initial evidence for such a discrepancy came from multiple studies on patients with neurological disorders (i.e., amnesia, short verbal span). Experiments on those populations gave two very important findings: first, they showed that there is a very clear distinction between short- and long-term memory, and second, at the same time there should be another type of structure which could not easily fall under either of those two categories (A. D. Baddeley & Warrington, 1970; Milner, 1966; Shallice & Warrington, 1970; Warrington & Weiskrantz, 1968; Wickelgren, 1968). In those studies, patients with damage in some areas of the temporal lobe showed a greatly impaired long-term memory as their performance in free-recall tests and questionnaires was very limited, whereas performance in tasks targeting the short-term memory capacity was completely unaffected (A. D. Baddeley & Warrington, 1970; Milner, 1966; Warrington & Weiskrantz, 1968; Wickelgren, 1968). Shallice and Warrington, 1970 testing subjects with conduction aphasia (repetition neurological defect with an extremely limited verbal span) found that performance in tasks targeting the short-term memory was impaired while in those targeting long-term memory their performance was normal. Following the proposed model by R. C. Atkinson and Shiffrin (1968), if short- and long-term memory use the same structures and functions, and long-term memory is an extension of short-term memory where items are solidified and stored indefinitely, then an impaired short-term memory might lead to problems with long-term memory which would be observable in the subjects' performance in long-term memory targeting tasks. That however was not the case suggesting that there must be components unique to short-term memory which function independently from long-term memory and deal with complex cognitive tasks.

Baddeley's working memory model

For these reasons, A. D. Baddeley and Graham (1974) proposed a model for working memory comprising of three different components, each one of them having a different role and being responsible for different functions. Instead of identifying that system as short-term memory, Baddeley and Hitch used the term working memory as it encapsulates better the idea that this part of the memory system is not simply one which stores and processes information simultaneously, but rather a subsystem that combines storage and processing by having specialised modules. The three main parts of that model included two buffers where information is stored (the phonological loop which deals with verbal information and the visuospatial sketchpad which is responsible for the visual information), along with a supervisory system mainly responsible for handling and processing the information stored in the other two com-

ponents, the central executive. An illustration of the proposed model is given in figure (1.5).

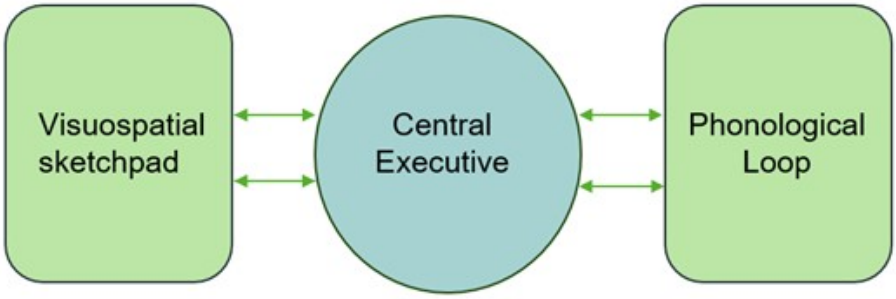


Figure 1.5: *The three-component model of working memory. This model, proposed by A. D. Baddeley and Graham, 1974, comprises a control system, the central executive, and two storage systems, the visuospatial sketchpad and the phonological loop.*

The phonological loop is the component that is responsible for briefly storing and processing speech-based information. It itself contains two further sub-components which work together, with the first one being a temporary acoustic or phonological store where memory traces are momentarily stored (for approximately two seconds) and an articulatory control process which has the capacity to refresh those memory traces before they fade by making use of subvocal rehearsal of those traces. Importantly, the phonological loop allows for information visually presented to be processed via the system, as long as this information can be phonologically or acoustically realised. For instance, if a subject visually sees and reads a word or even a string or a list of letters that can be processed acoustically, then those memory traces will be stored and initially processed within this system. Support for the existence of such a component as part of working memory has come from multiple studies which have suggested that word length has an effect on the speed of the subvocal rehearsal, with longer words delaying the process of refreshing the traces of the phonological information stored (i.e., A. Baddeley et al., 2002; A. D. Baddeley et al., 1975; Cowan et al., 1992), subvocal rehearsal is not the same with overt articulation as even subjects who do not have the capacity to articulate phonetically either due to age or due to muscular system issues still show evidence for subvocal articulation (i.e., A. Baddeley and Wilson, 1985; Bishop and Robson, 1989) and phonological similarity affects immediate recall, with phonologically similar words being significantly more difficult to recall as their traces are closely related and therefore more difficult to retain (i.e., A. D. Baddeley, 1966b; Conrad and Hull, 1964).

The visuospatial sketchpad similar to the phonological loop, is the system that holds visual and spatial information which it then manipulates and prepares for pro-

cessing. There have been some suggestions that the visual sketchpad much like the phonological loop might actually involve two separate sub-components as well, one which would be responsible for retaining visual information and a separate one for spatial information. Partial support to such claims has come from a rather limited number of studies testing mainly subjects with brain damage that have showed impaired performance in tasks targeting either one or the other while maintaining a generally unaffected alternative (i.e., Farah et al., 1988; Hanley et al., 1991), however most studies have treated those as quite similar properties that are processed by the same system rather than from different ones. The memory traces retained by this sub-system unlike to the phonological information stored by the phonological loop, have been shown to be based on multiple dimensions of a potential visual representation, with several studies finding different effects in performance based on stored cues of colour, shape, kinaesthetic, motor or spatial visual ones (i.e., A. D. Baddeley and Lieberman, 2017; Brandimonte et al., 1992; Klauer and Zhao, 2004; Logie, 1986; Nissen, 1985; Smyth and Pendleton, 1990; Smyth and Scholey, 1992).

The difference in the information that those two sub-systems, the phonological loop and the visuospatial sketchpad, store and process is evident from neurophysiological studies which suggest that lesions which affect processing and performance controlled by one of those domains might have no effect in processing cues by the other component. Clinical patients with lesions that might affect performance in tasks that target one's visuospatial processing, seem to perform similarly to normal subjects in tasks that target involving verbal processing and vice versa (i.e., Basso et al., 1982; De Renzi and Nichelli, 1975; Vallar and Baddeley, 1984a, 1984b).

Finally, the central executive is a supervisory system responsible for the distribution of the incoming sensory information to the appropriate subsystem, as well as controlling attention and dealing with other cognitive tasks. This component is much like what was proposed by Norman and Shallice (1986) as the *Supervisory Attentional System*, a system that partners with automatic patterns of response and together modulate everyday behaviour, by taking over control when automatic processing is not adequate. The central executive is described by A. D. Baddeley and Graham (1974) as the operator which not only allocates sensory cues to the other two components of the working memory system but it also connects them with long term memory, while at the same time it is the component that enables for the prioritisation of the incoming information in that it decides what should be processed, in what order and by which sub-system in the case there is a simultaneous flow of multiple stimuli (i.e., A. Baddeley, 2002; Shallice et al., 2002). The existence of a working memory component that is primarily a set of control processes rather than a storage and processing specialised one, comes from research which has shown that patients with degenerative diseases (i.e., Alzheimer's disease) while able to perform well in tasks that rely

on processing either by the phonological loop or by the visuospatial sketchpad, when asked to perform in a task that would require to some extent to combine those two processes they then show a performance greatly impaired (i.e., A. Baddeley et al., 1986; A. D. Baddeley et al., 1991; Logie et al., 2004).

So having as a starting point research that shows that in memory recall paradigms participants tend to quickly forget memorised objects or words unless those were continuously rehearsed or they otherwise formed a rather shorter than longer list (A. D. Baddeley, 1966a; Conrad, 1964; Murray, 1967; Wickelgren, 1965), A. D. Baddeley and Graham (1974) attempted to offer an explanation to such findings via the proposed model. Two of the most important ideas that A. D. Baddeley and Graham (1974)'s model appears to not only simply consider but to actually assign a central role to, are that the system has an extremely limited capacity and that the information that is processed gets activated and decayed; two concepts that importantly determine the way the model functions. In the model, the limited capacity of working memory, whether that is temporal (i.e., of around two seconds A. Baddeley (1992)) or number of individual chunks of informational cues, is actively shaping which part or which elements of the incoming information are temporarily held or immediately discarded. Working memory's limited capacity becomes evident as reaching full capacity in either of the sub-systems (i.e., phonological loop, visuospatial sketchpad) has an effect on the successful processing of incoming information by that specific sub-system. Evidently though, since they are two individual buffers, reaching full capacity in one of them does not have the same effect on the processing of information cues by the other sub-component. More specifically, if full storage capacity is reached for the storage of verbal information, then processing becomes disrupted, whereas if capacity is reached for visuospatial cues then processing of verbal cues is not affected (i.e., Farmer et al., 1986; Logie et al., 1990).

Overall then, the initial A. D. Baddeley and Graham (1974) working memory model as well as its further extensions (i.e., A. Baddeley, 2003) were an attempt to describe as accurately as possible the underlying structure and the distinct operations that need to take place in order to access temporary information as part of dealing with complex cognitive tasks. Since working memory is the system that deals with such tasks, it becomes rather clear that it must play an important role in sentence processing as well, as it impacts processes related to cognition, a system language is part of.

1.3.2 Working memory and Language

Indeed, multiple studies have shown that working memory and sentence processing are intrinsically linked. Depending on working memory capacity, which is widely

accepted that is limited by default (i.e., A. D. Baddeley et al., 1975, P. A. Carpenter et al., 1990, Engle et al., 1992, Page and Norris, 1998, Cowan, 2001, Saito and Miyake, 2004, Cowan, 2012), participants even based on their own individual capacities appear to have different patterns when performing in tasks in several experiments that target the processing of sentences (i.e., Daneman and Merikle, 1996, Rosen and Engle, 1997, Nicenboim et al., 2015, Von der Malsburg and Vasishth, 2013).

Working memory capacity has for long been linked with speed and comprehension, so the time participants require to read sentences in order to perform in relevant experiments, as well as whether participants are accurate in their judgement of sentences they have seen during experiments (Daneman and Merikle, 1996). Depending on the nature of the experimental task as well as the phenomena that participants were being tested on, working memory capacity span has been shown to affect behaviour in both ways; either it is that low working memory capacity leads to faster response or reading times or it is high working memory capacity that leads to such effects. The results on offline accuracy however, have been pretty consistent, with a clear direction where participants with a higher working memory span showing overall higher accuracy offline, suggesting that they have better comprehension of the experimental sentences they have encountered. Nicenboim et al. (2015) showed that differences in participants' working memory modulates their behaviour when tested on long distance dependency resolution, with participants with a higher working memory span showing more regressions and therefore take longer in the task than participants with a lower working memory span. Von der Malsburg and Vasishth (2013) also suggested that subjects' working memory capacity affects their processing of ambiguous sentences, with participants with lower working memory capacity having overall faster reading times but lower accuracy compared to the ones with high working memory capacity, a finding that has been observed in many other studies (i.e., Daneman and Merikle, 1996; MacDonald et al., 1992; Pearlmutter and MacDonald, 1995). Kemper et al. (2004) on the other hand, showed that when processing garden path sentences, a higher working memory span allows participants to adopt a processing strategy different from participants with a low working memory capacity especially at the point of disambiguation in the sentence, with high working memory capacity participants being faster than the low working memory capacity ones adapting their readings, as well as better in their accuracy offline. That was manifested in faster fixations for high span individuals and more regressions for low working memory span participants; so the difference in their behaviour based on working memory span was manifested primarily in fixations in later measures (i.e., regressions, total fixations time). In either case there seems to be a distinction between individuals with a high working memory span and the ones with a low span, whether it is offline where a high span gives rise to better comprehension, or online where it could

lead to different observations depending on the task, the timing and the phenomenon in question; the overall suggestion then is that irrespective of its direction, working memory capacity appears to modulate behaviour in linguistic experiments.

Having looked into how closely linked working memory and linguistic processing are, I will now present a class of models that have attempted to integrate memory architecture into sentence processing and parsing. Those models heavily rely on encoding and retrieving information based on individual features.

1.4 Cue-based Retrieval

Many researchers have attempted to incorporate the knowledge that generally the role of memory is vital for language and even more specifically that it is working memory that has the greatest and more targeted impact in its processing, by building language working models centred around the way that memory works. Those models have adopted the general mechanism of heavily relying on the cues that each item / word that is encountered is stored as, which then needs to be retrieved from memory to allow for the formation of dependencies between words that are non-adjacent. Multiple implementations have been proposed that follow such an architecture (i.e., McElree, 2000, Van Dyke and Lewis, 2003, McElree et al., 2003, Lewis et al., 2006, McElree, 2006, Van Dyke and McElree, 2011), but here I will focus on two of the most significant ones, which are simultaneously the ones presenting mechanisms that the studies that form the basis of this current investigation discuss; the activation-based model (Lewis and Vasishth, 2005) and the direct access model (McElree, 2000).

1.4.1 Activation-based model

One of the most central and still very influential models that attempted to bring together the knowledge on how memory works and tie it together with findings from the sentence processing literature is the language model presented in Lewis and Vasishth, 2005. Lewis and Vasishth built this model by employing the concepts of the limited capacity of working memory that Baddeley presented in his work (A. D. Baddeley and Graham, 1974, A. Baddeley, 2003), as well as the idea that items are activated and eventually decayed in memory, and formulated that computationally by making use of an Adaptive Control of Thought - Rational (ACT-R) architecture. What is suggested is that this type of framework is used when processing sentences, a framework that heavily relies on the working memory processes or activation and retrieval.

The basis of the Lewis and Vasishth framework is the ACT-R architecture, a theory that has been originally realised and further expanded in Anderson, 2013a, 2013b; Anderson et al., 1996, 1998; van Rijn and R Anderson, 2003. The ACT-R model investigates the role of working memory in a wide variety of cognitive phenomena, such as general problem solving situations (i.e., solving an equation, lexical decision, memory span when recalling digits), by taking into account capacity limitations and therefore the different stages of activation, decay, retaining and retrieval of information, while simultaneously bringing in rules that are in place for production and how those inform those memory processes. And eventually such an ACT-R model makes it possible to determine the probability and accuracy rate of an item being processed in memory, by utilising both production rules and the differing levels of activation of memory traces between different items.

This architecture was the starting point that made it possible for Lewis and Vasishth to bring together concepts from the domain of memory and more generally from theories of cognition with the domain of language processing, and utilise such a model for the purposes of issues that arise in sentence parsing.

According to Lewis and Vasishth, 2005, the resolution of dependencies, a phenomenon that has an exceptionally central role in language, and in particular dependencies that involve non adjacent elements (i.e., long distance dependencies) is only possible due a mechanism which retrieves those elements that have been encountered at some point within the sentence from memory. And more importantly, it does so based on their individual features that have been stored in memory when those elements were originally encountered.

As such, during the parsing of sentences each individual element / word is stored for a short period of time mainly in declarative memory as a chunk of separate cues (i.e., lexical, morpho-syntactic, semantic, structural, etc...) which are individually weighted.

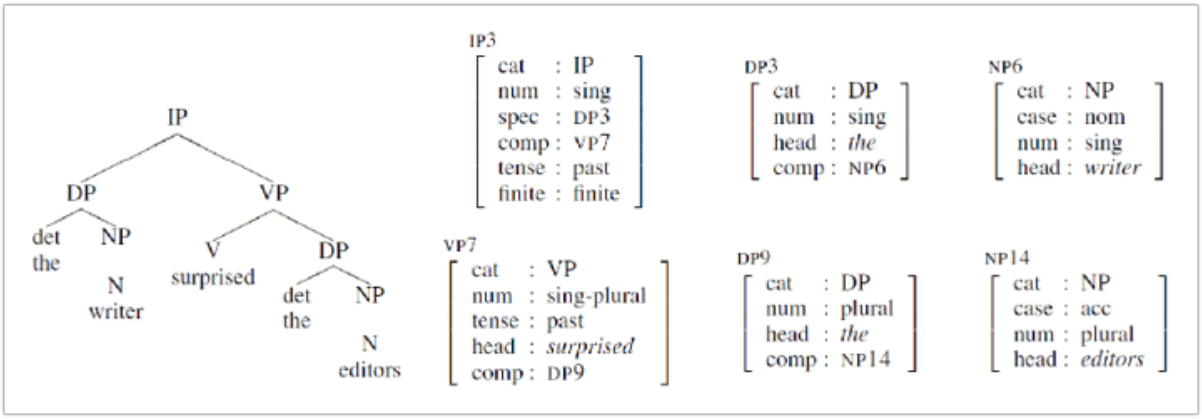


Figure 1.6: *Example of how a whole sentence is potentially stored based on features.*

As illustrated in figure 1.6, for each of the elements in the sentence *the writers surprised the editors*, a chunk of features is generated, with features in those being the category, the number, the tense, etc..

The present model by Lewis and Vasishth follows a line of research that assumes that the encoding of new incoming information in memory is possible due to those new elements being stored as individual chunks of specific features (Miller, 1956; Nairne, 1990; Oberauer & Kliegl, 2006; Servan-Schreiber, 1991). As such it supports the idea that upon encountering each word in sentence in figure 1.6, a chunk of separate features will be generated and eventually encoded (stored for a short period of time), until it is needed in processing, so until either its retrieval (being brought back to focus)or its eventual decay.

Nairne (1990) assume that memory traces of new elements are encoded as vectors that contain different types of features that characterise that element, while those features are also assigned an individual weight. As shown in the simulations for memory recalls that Nairne, 1990 ran, the probability of recalling and recovering an item from memory is dependent on those stored features. At the same time, even frameworks that adopt a different approach to modeling (i.e., Oberauer and Kliegl (2006) fitting a mathematical model to language data by utilising non-linear mixed effects models (NLME)) also assume that the interaction with items in memory is only possible due to each item being stored as a bunch of features that will eventually be utilised to bring them back to focus. Such vectors of feature representations of individual items in memory have also been researched by Lange and Oberauer (2005) and Neath (2000) among others, suggesting that when it comes to the retrieval from memory of items necessary for processing in order to complete a linguistic task, the accuracy and the time needed are bound to an item's stored representations of its individual features.

The model presented in Lewis and Vasishth (2005) assumes that the main factor that determines which chunk gets activated is the level of activation that it receives. When it comes to the resolution of a dependency -so in cases where the retrieval of chunks will be essential- the level of activation that a chunk receives will be based on a match between the feature-cues at the point of retrieval and the features-cues that it contains. For the successful resolution of a dependency, the chunk that is retrieved has to have a high degree of match between its cues and the cues at the retrieval point, which will subsequently lead to it receiving a higher level of activation compared to any other chunks stored in memory. The total sum of the levels of activation that each element receives, is the combination of the base activation of the chunk which is dependent on its history of past usage as well as the frequency of usage, so how recently the particular chunk was retrieved as well as how often, and an activation boost that the chunk receives based on the match between the

cues that the element's chunk contains and the cues that are at the retrieval probe modulated by the individual weight associated with each of those cues. At the same time, the lack of a match between one or more of the feature-cues of an element and the ones at the retrieval probe of course always depending on their individual weights, is penalised in such a way that less activation is received overall, eventually leading to that particular element's memory chunk faster decaying in declarative memory. Overall then, the total activation of an individual chunk (combination of base activation, together with the sum of match boost or mismatch penalty activation) essentially determines both the probability of its retrieval as well as the latency of that retrieval.

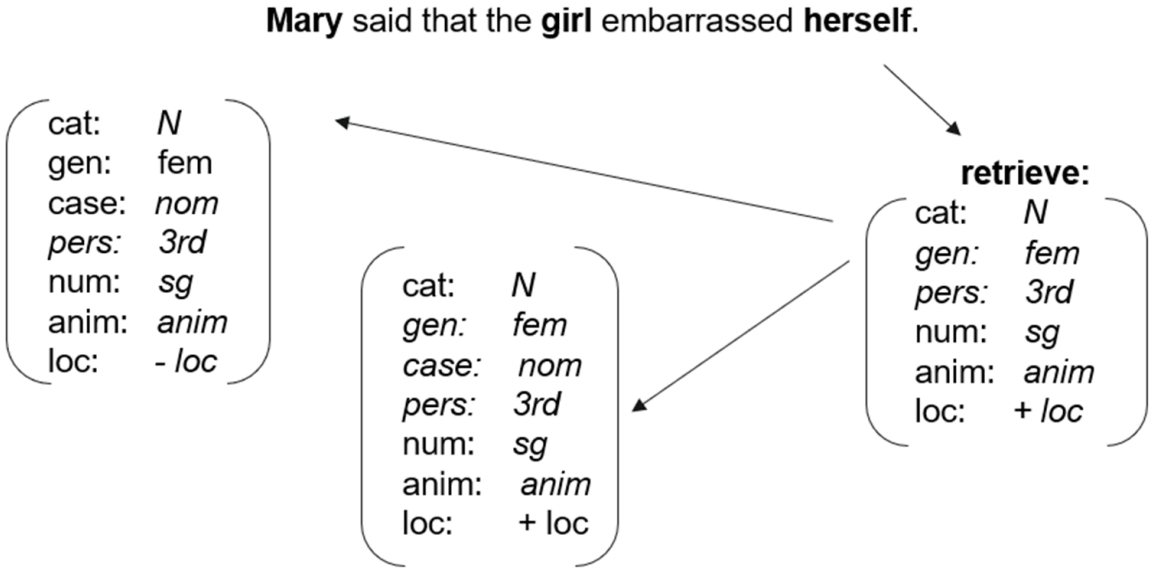


Figure 1.7: *Example of how a whole sentence is potentially stored based on features.*

For instance, upon encountering sentences like the ones in figures 1.7 and 1.8, the Lewis and Vasishth (2005) model makes some very clear predictions.

In the sentence in 1.8 the parser stores information on each different element (i.e., on the word *Mary* and on the word *boy*). Upon encountering the word *himself*, the parser realises that it needs to retrieve a previously encoded item so that it resolves the dependency. The chunks of features for both of those words (*Mary, boy*) do contain matching features with the ones at the retrieval probe, for example the category, the number and the animacy features, so both of those chunks are expected to receive a level of activation. However, the word *boy* also shares the features locality as well as the feature gender with the features at the retrieval probe, and therefore it is expected to receive an activation boost. On the other hand, the word *Mary*, has some stored feature-cues that are not a good match for the retrieval probe, such as locality and gender, and is therefore expected to be penalised by receiving lower levels of

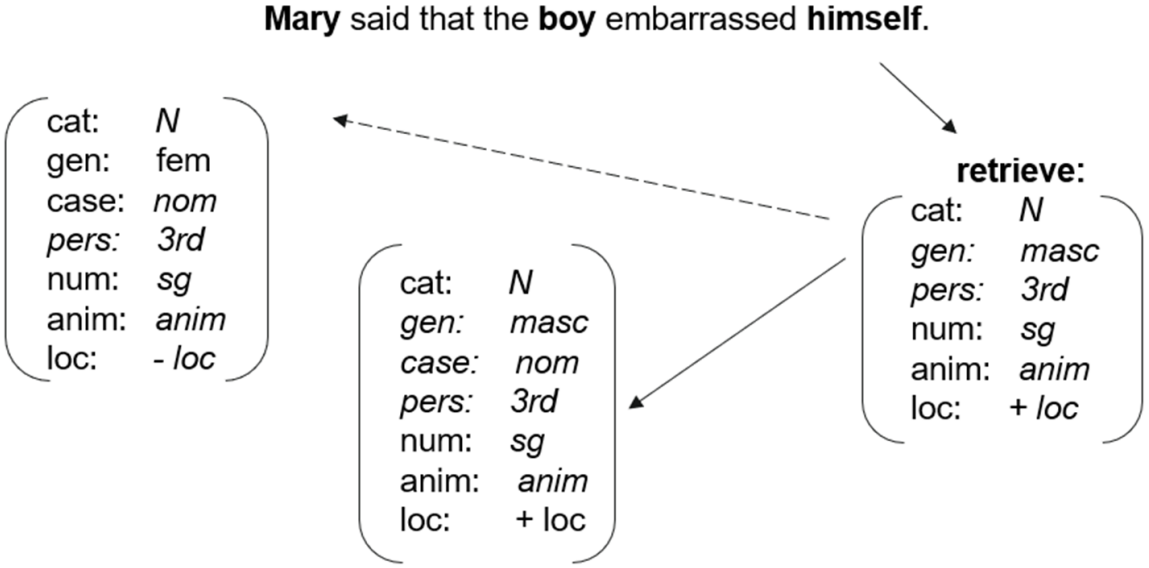


Figure 1.8: *Example of how a whole sentence is potentially stored based on features.*

activation. Eventually, the cue-based model predicts that what will be eventually retrieved will be the chunk of features for the word *boy* in the sentence in figure 1.8, leading to an integration of the element *boy* with the word *himself*.

On the other hand, in the sentence in figure 1.7, the chunks of features for both words (*Mary, girl*) are expected to receive higher levels of activation overall, as they both contain features that highly match those at the retrieval probe. Therefore it is expected to be harder for the processor to establish which element should be retrieved and integrated into the parse. If comparing the parsing of the two sentences then, the Lewis and Vasishth (2005) cue-based model predicts that the processing of *herself* in the sentence in figure 1.7 should be more difficult than the processing of the word *himself* in the sentence in figure 1.8.

Importantly however, what is not defined by the cue-based model is that there is no limitation as to what constitutes a feature for an element and neither is clear how exactly those features receive their individual different weights.

Therefore, memory processes appear to be the focal point in the Lewis and Vasishth framework (Lewis and Vasishth, 2005), as it is those memory processes that allow for accurate sentence parsing through the storage, processing and retrieval of those chunks of feature-cues. The model differentiates between different types of memory (working, declarative, procedural) which when working together enable the successful parsing of sentences, while it simultaneously highlights the important role of working memory. Essentially, what it is suggested is that in this framework, the system functions in such a way, where the chunks including the different sets of

feature bundles associated with each encountered element within the sentences are stored primarily in declarative memory, with the only exception being the element currently attended and one extra element, both of which are processed by working memory. Based on this framework, when attempting the resolution of a dependency between non-adjacent elements, the parser at the encountering of the cues at a retrieval probe (so at the point where it needs to resolve the dependency to successfully make sense either locally or of a whole sentence) is required to retrieve a chunk associated to an element previously encountered. Having the retrieval cues at the retrieval probe in working memory (as it is the currently attended input and therefore at the focus of attention) it needs to resolve the dependency by searching and eventually retrieving and bringing back to focus another chunk which includes feature-cues that are a good match for the feature-cues of the retrieval probe. The search and subsequent retrieval of a chunk has to be through the declarative memory as in a long distance dependency, one of the elements of the dependency would have been encountered earlier in the sentence and therefore the associated chunk of features would not be currently in the focus of attention. The retrieval of a chunk of feature-cues that has been stored in declarative memory is mediated through the procedural memory, which enables that chunk to be brought back to focus and therefore become available to working memory so that it can be further processed and subsequently lead to the resolution of the dependency in question.

(21) a. **Mary** said that the **girl** embarrassed **herself**.

b. **Mary** said that the **boy** embarrassed **himself**.

Dependency resolution according to the cue-based model then works as such: in the case of the two sentences above shown in figures 1.8 and 1.7 (also included in example 21), the cue-based model assumes that first the chunks of features for the elements *Mary, girl / boy* are stored in declarative memory. Upon encountering the word *himself / herself* which is the input currently attended in working memory and detects the retrieval probe features, the parser searches back in the declarative memory mediated by the procedural memory, to find chunks of fitting features associated with elements previously attended. When it detects the most fitting feature bundle, it receives the highest level of activation and it is brought back to focus, becoming available in working memory. There, the specific element associated with the chunk of features that got the highest level of activation can be processed. When both elements are in focus then (i.e., *boy* and *himself* in the case of sentence b. in the example in 21), they can be processed so that a dependency gets established.

Interestingly, when it comes to ungrammatical sentences, the cue-based model can make predictions of facilitation in processing. For instance, in an example like 14 from Bock and Miller, 1991 (*The key to the cabinets are...*), the cue-based model

assumes that chunks for each one of those words would be stored in memory. Upon encountering the verb in plural *are*, the parser would look for that feature in the chunks of other items in the sentence. Considering that the plurality feature is not found in the chunk of the grammatical subject (namely the *key*), the parser finds it in the chunk of the *cabinets*. Despite the sentence then being ungrammatical altogether, the processing becomes facilitated compare to the processing of a sentence that is equally ungrammatical (i.e., *The key to the cabinet are...*) but with no partial match between the features of the verb (*are*) and any other item.

An important aspect of the Lewis and Vasishth model and one that is the determining factor for the predictions that the model makes is that the retrieval of an element from memory is considered to be the result of a race, a process that is similar to what other models have also suggested that is the driving force behind processing (Audley & Pike, 1965; S. D. Brown & Heathcote, 2008; Logačev & Vasishth, 2016; Ratcliff & McKoon, 2008; Usher & McClelland, 2001; Van Gompel et al., 2001; Van Maanen et al., 2012; Vickers, 1970). When the parser needs to retrieve an item from declarative memory, the item that will be eventually retrieved will be the one that *wins* the race among the competitor elements while the time that will be needed for its retrieval will be dependent on how quickly that element has accumulated the highest levels of activation among the other stored items. What the model highly prioritises is not simply the high levels of activation but at the same time the speed with which those get allocated to a bundle of features. Overall then, in the Lewis and Vasishth framework, the process of retrieval which is what allows for successful sentence parsing is in itself a process that is driven by a race of the fastest activation between different elements.

1.4.2 Direct access model

Another cue-based model, slightly different to the one proposed by Lewis and Vasishth (2005) is the direct access model (McElree, 2000).

While the direct model and the activation based model are usually used interchangeably, they are based on different assumptions and make different predictions. Essentially, they predict differences in the relationship between reading times and response accuracy. The activation-based model as discussed earlier assumes that the retrieval process is mainly a race between accumulators of increased activation, with the accuracy of the retrieval being determined by the winner of the race and retrieval time by its rate of activation accumulation. In contrast, the direct-access model assumes that only the probability of retrieval can be affected, while the retrieval time is a result of the possibility of backtracking and repairing incorrect retrievals.

The direct access model agrees with the activation based model that the probabil-

ity of retrieval for any item from memory given the retrieval features is a function of the degree of the match between the features and the degree to which those feature match other competitor items in memory. However, in contrast with the activation-based account, features are considered to enable direct access to relevant memory representations.

Those models will be collectively referred to as cue-based models in the rest of the thesis as they do not make very different predictions on the dependencies in question.

An effect that those models can better explain through the adoption of cue-based mechanisms is interference in processing.

1.5 Similarity-based Interference

Based on the predictions and general architecture of either of the presented models (Lewis & Vasishth, 2005; McElree, 2000) as well as any other models that heavily rely on cue-based memory retrieval (Lewis et al., 2006; McElree et al., 2003; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2011) is interference between the different elements and the features they share when stored in memory (i.e., Jonides et al., 2008). The concept of elements being stored as chunks in the declarative memory where they are retrieved from via competition, brings up the question of what might happen if chunks of features stored in memory potentially share some of their cues. In those cases, especially when certain elements that belong in the same sentence share some of their cues in their stored chunks, similarity based interference arises, as in that there is more competition when selecting for activation (bringing back to focus) between the different chunks and therefore the associated linguistic items. The parser then behaves differently, either in processing / retrieval times or accuracy in retrieving the correct item. Even more so, in cases where the parser has to resolve dependencies, and in particular when those dependencies involve non-adjacent elements, some similarity between the cues of the structurally accessible element and the stored features of any other inaccessible elements (distractor words) might make the parser's task even more demanding.

As seen earlier in figure 1.7, the cue-based model predicted that processing the dependency would be harder, due to both the word *Mary* and the word *girl* sharing lots of features with the retrieval probe features at *herself*. That is a case of similarity based interference, as the similarity between the features of those two elements (i.e., *Mary*, *girl*) is expected to interfere with the dependency's processing.

1.5.1 Proactive vs. Retroactive Similarity-based Interference

When it comes to interference phenomena rising as a result of the feature similarity between cues, Lewis (1996) first introduced the terms of proactive versus retroactive interference in sentence processing. Essentially, those phenomena are observed when in the process of establishing dependencies, the parser encounters items that are not part of the dependency (distractor items), but which have the potential of affecting the dependency's resolution or general processing due to those distractor items sharing some of their feature / cues with the dependency's target item when stored in memory.

Those two different types of interference have been established in the literature depending on the position of the material which does not belong in a particular dependency in a sentence is presented in relation to the items that form the dependency itself. When the distractor material intervenes between the presentation of two constituents that form a dependency then it is a case of retroactive interference. The presence of such a type of material has been shown to have an effect on the encoding, the storage or even the retrieval of the correct items that form the dependency (Cunnings & Sturt, 2018; Van Dyke, 2007; Van Dyke & Lewis, 2003; Wagers et al., 2009). In a sentence like: *Sue remembered the plate that the butler with the cup accidentally shattered today in the dining room.*, the distractor word *cup* intervenes between the retrieval launch site *shattered* and the target word *plate*.

When structures appear before the presentation of material that form a constituency then it is a case of proactive interference. Presenting potentially relevant, whether that could be semantically or syntactically, fitting material before the presentation of the actual constituents, might also affect the processing of the dependency as it might lead to delays, longer reading times or general increased difficulty in processing of a particular dependency, either during the encoding, the storage or the retrieval stages (Cunnings & Sturt, 2014; Fedorenko et al., 2006; Gordon et al., 2002; Van Dyke & McElree, 2006, 2011). For instance, in a sentence like: *It was the boat that the guy who lived by the sea fixed in two sunny days.* which is preceded by the three words table-sink-truck, the distractor items *table-sink-truck* precede both the retrieval launch site *fixed*, and the target word *boat*.

For the cases of retroactive interference, the previously discussed cue-based models would predict that the stored features of potentially relevant representations might interfere with the stored features of an item involved in a dependency. The representation of the item that has been presented earlier must be maintained in memory for longer, so naturally it might decay over the time that is needed before encountering the other part of a dependency or the features that have been stored might be displaced from memory from the features of the intervening relevant material. That

would then mean that the processing of the dependency at any stage during processing might be affected, as the potential similarity between the features of one part of a dependency and the features of the intervening material, as well as the natural decay in memory due to the distance between the material of the dependency might lead to interference.

Regarding proactive interference the cue-based models would predict that if the representations of structures, that have been stored in memory prior to encountering items that form a dependency, contain features that partially match those at the retrieval probe of the dependency, then the processing of that dependency should pose increased difficulty compared to the processing of dependencies not preceded by partially matching material. The cue-based models suggest that if multiple similar items are stored in memory then it might lead to interference either during the encoding of one of the items of the dependency (i.e., longer reading times / duration of fixations upon encountering an item that might be similar to something previously stored or / and accuracy in retrieval, depending on the particular model), during the storage or at the retrieval stage of the item of the dependency (i.e., longer reading times / duration of fixations at the retrieval point in the structure, or / and accuracy, again depending on the model).

1.5.2 Facilitatory vs. Inhibitory Similarity-based Interference

Another aspect of cue-based similarity between different items that has been widely discussed and researched is the effects that it has on the time that subjects need when processing the structures that contain them (i.e., (Badecker & Straub, 2002; Jäger et al., 2015; Pearlmutter et al., 1999; Van Dyke & McElree, 2006), just to mention a few).

When an intervening distractor item that is not part of a certain dependency in a sentence but shares a lot of its feature / cues with the cues at the retrieval point of that dependency, or in some cases might even be a better match to that retrieval probe than what is grammatically correct based on the material in the sentence (i.e., in ungrammatical sentences) and leads to the attenuation of the reading times which do not however always reflect processing times for restoring the dependency, then that is a case of facilitatory interference. On the opposite, when the existence of one or more potential distractor items which share feature / cues with the retrieval probe in a dependency lead to a slowdown manifested in overall slower reading times, then that is what would be considered as inhibitory interference.

Importantly, those two different kinds of interference can be predicted and to a point explained by the activation-based model and the direct access model in a partially similar but nevertheless distinct fashion.

Inhibitory interference is a phenomenon that is differently explained by the two

models presented above (Lewis & Vasishth, 2005; McElree, 2000). The Lewis and Vasishth (2005) activation-based model would describe inhibition occurring in sentences where there more than one potential matches to the retrieval probe (so at least one distractor item competing with the target item), as the direct result of the share of activation between all the potential items that share some feature / cues with the cues at the retrieval site. The ACT-R model assumes that the activation due to matching the cues at the retrieval site will spread to all potential items, and therefore the amount of activation to the correct target item will be reduced. As a result, the latency of the retrieval of the grammatically correct target item will be increased leading to delays in processing, while the probability of misretrieving one of the grammatically unlicensed distractor items is also higher. However, the direct access model and its further implementations (McElree, 2000; McElree et al., 2003) have an overall different approach to inhibitory interference phenomena. The direct access model puts a lot of emphasis on the probability of retrieving the correct target item, while the speed of its retrieval is not directly affected. Based on this model, the existence of more than one potential matches (sharing some of their feature / cues with the ones at the retrieval probe) leads to an overall decreased probability of retrieving the correct target item to resolve a dependency. In sentences where there are more than one partially matching items, the model predicts that the probability of retrieving the correct target item is lower, while the probability of retrieving any of the competing distractor items is higher. The associated delay that is manifested in increased reading times then is explained as a result of the higher probability of error, so of potentially retrieving the wrong item (one of the distractor items) due to its raised probabilities of retrieval, and therefore the associated repairing in order to process the structure correctly. Such a process would involve the parser wrongly retrieving one of the distractor items, then backtracking in order to reprocess the dependency and eventually reaching the retrieval of the correct item. Therefore, the latency that is associated with inhibitory interference is not directly associated with the accessing and retrieving the items for processing a dependency, but rather the parser's error making and error fixing due to the similarity of probabilities of each of the items competing for retrieval.

Facilitatory interference however, is approached differently by the two models previously discussed. Based on the activation model by Lewis and Vasishth (2005), any match between the cues at the retrieval probe and the feature / cues of any of the potential items (target and / or distractors) is subsidised, such that it is given an activation boost. The model assumes that there is indeed a race between multiple items whose features match those at a retrieval probe, so since all of those compete with each other for retrieval, the average reading times are faster than in cases where there is no competition between different items (so in scenarios where a target item is the only one sharing multiple features with the retrieval probe, whereas other

items in the sentence would not be considered as a good match to any of those said features / cues, reading times would not be expected to be as fast, as there is no race process). Nevertheless, the direct access model does not predict such a phenomenon like facilitatory interference. It would not predict a difference between the probability of retrieving any of the potential items to lead to faster overall reading times. The direct access model, primarily seems to treat differences in reading times as a direct result of reanalysis which is itself a result of misretrievals. The direct access model then cannot necessarily explain the potential speedup in reading times as an interference effect, as the latency of the retrieval of different items from memory is not directly affected by that item's probability of retrieval in a given sentence.

1.5.3 Interference in stages: encoding vs. retrieval

Importantly, some of the studies previously discussed report reliable online effects, such as longer reading times, that are taken to suggest that interference is at play complicating the processing of the discussed structures. The reported online reading times though are of the region that includes the second noun phrase in a given sentence (Fedorenko et al., 2006; Gordon et al., 2002) which matches with the class of the memorised nouns. For instance, in the example 12 from Gordon et al., 2002, longer reading times were found at the region containing the noun *fireman* in 12b.b. when preceded by the three memorised nouns (i.e., *poet - cartoonist - voter*), compared to the region with the same word (i.e., *fireman*) in example 12a.b. where it is preceded by three memorised proper names (those would be *Joel - Greg - Andy* in the example). Similarly, in example 13 longer reading times were found for the region containing the second noun (i.e., *cardiologist* in example 13) when preceded by a memory load with three other similar nouns (that would be occupation nouns in the example). However, the same regions in both those studies also include the verb of the sentence and therefore the critical dependency (those would be the verb *liked* in example 12 and the verb *consulted* in example13). Consequently, it is not clear at what stage interference affects processing and subsequently leads to longer reading times. In other studies despite the consistent interference effects, it is also difficult to tease apart at which particular stage during processing they arise (Gordon et al., 2001; Jäger et al., 2015; Laurinavichyute et al., 2017).

Based on a general cue-based model architecture, interference could occur at any point during processing, either during the encoding of incoming information or at the stage of retrieval, so when the parser tried to establish a relationship between different elements. However, identifying the point when interference has an effect, especially when discussing sentence external distractor items should be central when discussing the processing of long-distance dependencies. If the interference effects

previously discussed are observed during the encoding stage of processing, then that would potentially suggest that when the parser has to store yet another item in memory that involves features very similar to the ones of other elements recently stored, it faces a difficulty. If however the interference effects in those studies are attributed to the retrieval stage, then that would imply that elements already stored in memory even if they do not belong in a particular sentence can affect the way long-distance dependencies are built inside a sentence.

For example, Gordon et al. (2006) replicated the results from Gordon et al. (2001) in an eye-tracking paradigm, testing object-extracted vs. subject-extracted relative clauses which involved two noun phrases, where the second could either match the first by being a description noun or be a proper name. An example of the experimental items is given in example 22.

(22) a. The banker that praised **the barber / Sophie** climbed the mountain just outside of town.

b. The banker that **the barber / Sophie** praised climbed the mountain just outside of town.

The results clearly showed that similarity based interference in the tested sentences occurs at the encoding or maintenance stage and not during retrieval. When there was a match between the two noun phrases involved in the object-extracted relative clauses (both being description nouns), the participants' fixations were significantly longer in the relative clause region (i.e., *that the barber praised*) during right-bounded reading, which is a relatively early measure of processing and a similar pattern of longer fixations was also reported during regression path duration on the matrix verb (i.e., *climbed*). It is suggested that these results can only be explained by similarity based interference at the encoding or maintenance stage, as it is during the early processes (manifested in the early eye-tracking measures) that a significant difference is observed, implying that the difficulty in processing stems from storing or maintaining the unique features of two similar noun phrases that are to be interpreted further down in the structure as arguments of the verb, and not necessarily due to the difficulty of retrieving them, as that would probably be manifested in later processing measures.

Results from a study in Italian also provide some evidence of similarity based interference effects at the stage of encoding. In a self-paced reading design, Villata et al. (2018) tested sentences like the ones in example 23 manipulating the gender of the object (masculine vs. feminine), and the match between the gender of the subject and the gender of object (match vs. mismatch), followed by comprehension questions that always targeted thematic role attribution in the relative clause (e.g., *Did the waiter surprise the dancer?* vs. *Did the dancer surprise the waiter?*).

(23) a. Il/ballerino/che/il/cameriere/ha/sorpreso/beveva/un/cocktail/alcolico
The/dancer-MASC/that/the/waiter-MASC/has/ surprised-Ø/drank/a/cocktail/with alcohol

b. Il/ballerino/che/la/cameriera/ha/sorpreso/beveva/un/cocktail/alcolico
The/dancer-MASC/that/the/waiter-FEM/has/surprised-Ø/drank/a/cocktail/with alcohol

c. La/ballerina/che/la/cameriera/ha/sorpreso/beveva/un/cocktail/alcolico
The/dancer-FEM/that/the/waiter-FEM/has/surprised-Ø/drank/a/cocktail/with alcohol

d. La/ballerina/che/il/cameriere/ha/sorpreso/beveva/un/cocktail/alcolico
The/dancer-FEM/that/the/waiter-MASC/has/surprised-Ø/drank/a/cocktail/with alcohol

Faster reading times were observed at the word *surprised* in sentences where there was a mismatch between the gender of subject and objects nouns so in sentences 23b. (containing *dancer*-MASC – *waiter*-FEM) and 23d. (containing the words *dancer*-FEM – *waiter*-MASC). Accuracy results were also consistent, with participants being more accurate when there was a mismatch of gender. Those findings were taken to be the result of encoding interference, as in Italian the past participle does not agree in gender with the subject, so gender could not be considered as a retrieval cue.

1.6 Retrieval interference effects in filler-gap dependencies

A dependency of a particular interest that falls under the category of long distance dependencies, is filler-gap dependencies. As previously discussed those are dependencies that by default involve an item which is displaced (a filler) that has to be retained in memory and eventually retrieved and associated with a gap, so that a sentence can be successfully parsed. Since the filler item has to be stored and retrieved, a cue-based framework would predict that if distractor items appear either before or after the filler item, they could potentially interfere with the resolution of the filler-gap dependency if they are a good match to the feature / cues at the gap (where the dependency is resolved) leading to changes in a sentence's parsing. As such the cue-based model could predict either facilitation in those cases or inhibition

in processing, depending on whether the stored features / cues of the filler item are a good match to the ones at the gap.

Two studies that have investigated semantic retrieval interference in filler-gap dependencies by utilising two different kinds of interference are the Van Dyke and McElree (2006) study focusing on proactive inhibitory interference effects and Cunnings and Sturt (2018) finding retroactive facilitatory interference effects.

More specifically, Van Dyke and McElree (2006) have shown that in filler-gap dependencies when some distractor items precede the critical filler word and they also match the cues at the retrieval gap, the processing of the filler-gap dependency becomes more challenging for the parser compared to the processing of the same dependency in the absence of matching preceding distractors. Importantly based on the results of the study, it is claimed that the distractor items do not even need to be within the sentence which contains the retrieval gap and therefore the critical dependency, but can still interfere with the processing of said dependency.

On the contrary, when distractor items that match the retrieval cues at the gap appear after the critical filler in a filler-gap dependency, processing becomes facilitated as shown in Cunnings and Sturt (2018).

1.6.1 Proactive Inhibitory Retrieval Interference

Van Dyke and McElree (2006) tested for retrieval effects caused by distractors outside of a sentence during the resolution of an intra-sentential filler-gap dependency. object-clefted sentences were tested in a dual task self-paced reading design, similar to the one employed by Gordon et al. (2002), manipulating the type of verb and the existence of an additional memorisation and recall task of three nouns (an example of the experimental items used is given in table 1.1).

$+/-$ **memory load:** table/sink/truck
Non-interfering: It was the boat that the guy who lived by the sea **sailed** in two sunny days.
Interfering: It was the boat that the guy who lived by the sea **fixed** in two sunny days

Table 1.1: Example experimental items in Van Dyke and McElree, 2006

In the memory load condition, participants were asked to read the three nouns out loud and memorise them, before proceeding to reading the sentence in a self-paced reading design phrase by phrase. The sentence was followed by a comprehension question, after which subjects were asked to recall the three memorised nouns and write them down in the order they had originally seen them. In the absence of a memory load, participants had to simply read the sentence phrase by phrase and answer the subsequent comprehension question. The manipulation of interference referred to the relationship between that verb and the three memorised nouns in the

memory load, such that in the interfering condition the three nouns could be potentially plausible arguments of the verb, whereas in the non-interfering they would be implausible arguments of the verb.

The most important finding in the study according to Van Dyke and McElree (2006) was a significant interaction of Interference and Memory Load ($p < .04$), such that at the region of the critical verb (i.e., *sailed / fixed*) participants were slower at the interfering condition when there was a memory load. That was taken to suggest that since the complexity of the two sentences is kept constant in the two conditions and does not lead to an effect (lack of an effect in the reading times in the no memory load condition), the only difference is the relationship between the three memorised nouns and the verb in the sentence. The found effect then could be attributed to that particular different relationship between the three nouns and the verb. That would be that the three nouns in the memory load interfere with the retrieval of the filler noun, as all four words (the three memorised nouns and the actual filler noun) would match many of the cues at the retrieval probe, leading to competition for retrieval from memory and subsequent delay in processing. Van Dyke and McElree (2006) suggested that these results are in line with the expectation of the cue-based model as outlined by Lewis and Vasishth (2005) and Van Dyke and Lewis (2003), which would expect the stored items in memory to receive activation as a result of the match between their cues and the cues at the retrieval probe (in this case, the verb).

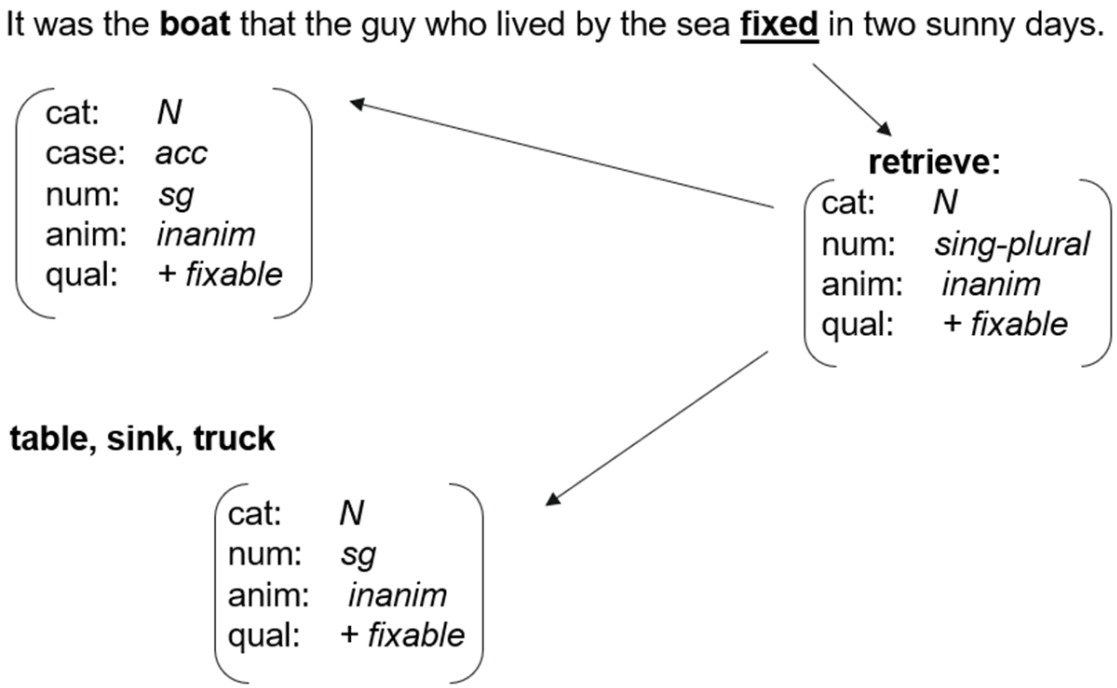


Figure 1.9: *Example of sentence in the interfering condition in Van Dyke and McElree (2006) as stored chunks of features.*

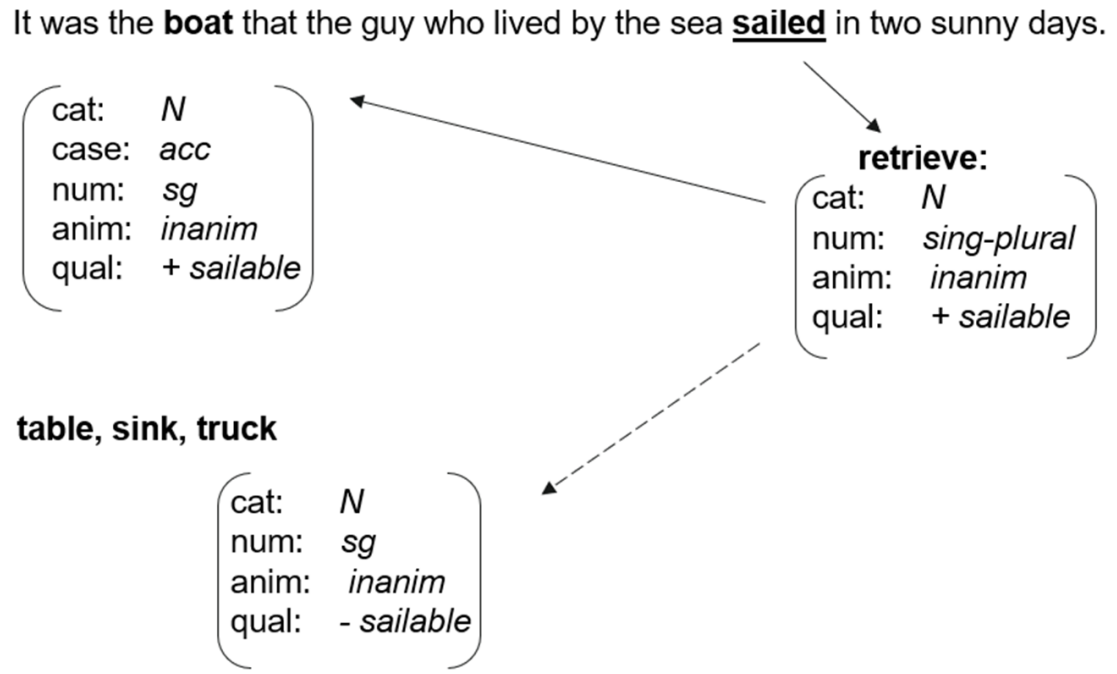


Figure 1.10: *Example of sentence in the non-interfering condition in Van Dyke and McElree (2006) as stored chunks of features.*

Using the same example sentence given in table 2.1, figure 1.9 shows that the three nouns in the memory load which the participants in the study were asked to memorise (i.e., *table, sink, truck*), are all *"fixable"* as is the grammatical filler of the verb (i.e., *boat*), but they are not *"sailable"* as seen in figure 1.10. They also share other features with the ones at the retrieval point, for example they are all nominal, and inanimate.

Therefore according to Van Dyke and McElree (2006), their results show that this *"fixable"* feature that all the four nouns share with the retrieval probe renders them potential retrieval targets, which leads to the observed interference effects and subsequently to longer reading times in the interfering condition compared to the non-interfering one. The absence of a similar effect in the non-interfering condition is attributed to the lack of a + *sailable* feature for the three memorised nouns, and therefore them not being a good match for the retrieval probe.

Overall then, the Van Dyke and McElree (2006) study claims that when the parser tries to resolve dependencies in a particular structure, other linguistic elements that are not necessarily in the sentence could potentially lead to interference and make parsing more difficult as they are considered as equally plausible alternatives for retrieval, an idea that is in line with the direct access model. That then means that in a cue-based framework, structural information is not considered to the same extent as semantic similarity when the parser tries to resolve filler-gap dependencies.

Van Dyke et al. (2014) attempted to replicate this interference effect. Using the same object clefted sentences as Van Dyke and McElree (2006) and with the same manipulations (manipulating the presence of a memory load and interference, so whether the words in the memory load could be potential plausible direct objects of the main verb in the sentence), they tested participants with varying levels of working memory capacity. Participants in the experiment were tested in a series of 24 pre-tests to identify their working memory capacity, with the expectation that the results of Van Dyke and McElree (2006) should be replicated at least for the high working memory capacity subjects. However, the expected online effect was not found in this study, regardless of the working memory capacity of the tested subjects.

Mertzen et al. (2020) ran a cross-linguistic study in English, German and Russian to test for similarity based retrieval interference effects from sentence external distractors during the resolution of within sentence dependencies. An eye-tracking experiment was run in all three languages, while the manipulations were similar with the ones in Van Dyke and McElree (2006) and Van Dyke et al. (2014), manipulating the presence or absence of a memory load and the interference between the memory load items and the main verb in the sentences. The additional manipulation in this study was the complexity of the comprehension questions, as they tried to see whether the expected similarity based retrieval interference effects would change de-

pending on the required depth of processing. Mertzen et al. (2020) expected to find larger interference effects when the comprehension questions required the correct resolution of the dependencies in the sentence, whereas the simpler questions which did not require resolution of dependencies were expected to lead to milder / more limited effects. Mertzen et al. (2020) found no effect in the total reading times in any of the tested languages and regardless of the complexity of the comprehension question manipulation.

From the studies presented here it is difficult to conclude whether sentence external distractors do lead to interference effects in the resolution of within sentence dependencies (filler gap dependencies in this case). The data provided seem to be confusing, as the effect is only observed in Van Dyke and McElree (2006), whereas despite the manipulations and the items being almost identical no such effect is observed neither in Van Dyke et al. (2014) nor in Mertzen et al. (2020).

1.6.2 Retroactive Facilitatory Retrieval Interference

Cunnings and Sturt (2018) also discuss interference effects when restoring dependencies in filler-gap sentences. They investigated memory retrieval interference during filler-gap dependency resolution as well, by utilising plausibility effects. Two of the main differences from the previously discussed studies (Mertzen et al., 2020; Van Dyke & McElree, 2006; Van Dyke et al., 2014), are that the focus here was on interference effects caused by sentence internal distractor nouns which partially match the verb's cues (retroactive interference), while the reported retrieval interference leads to facilitation in processing rather than inhibition as is the case in Van Dyke and McElree (2006).

Sentences which involved a retrieval target noun, as well as a distractor noun which could potentially interfere with the resolution of the filler-gap dependency depending on its plausibility were tested in an eye-tracking design. The main manipulations were the plausibility of the sentence as well as the plausibility of the a distractor noun, intervening between the filler noun and the critical verb of the sentence (example of experimental item in table 1.2). The plausibility of the sentence was manipulated such that the grammatical object was either a plausible or an implausible argument of the verb in the relative clause (i.e., *shattered... the plate* vs. *shattered... the letter*). The plausibility of the distractor noun was also manipulated, so that it was either a plausible or an implausible argument of the same verb (i.e., *shattered... the cup* vs. *shattered... the tie*).

Since plausibility is what is being manipulated in this study, the found effects are being discussed as illusions of plausibility, where seemingly plausible distractors might interfere with the resolution of dependencies and therefore the comprehension

Sentence	Distractor	Sentence
plausible	*plausible*	Sue remembered the **plate** that the butler with the **cup** accidently shattered today in the dining room.
	implausible	Sue remembered the **plate** that the butler with the **tie** accidently shattered today in the dining room.
implausible	*plausible*	Sue remembered the **letter** that the butler with the **cup** accidently shattered today in the dining room.
	implausible	Sue remembered the **letter** that the butler with the **tie** accidently shattered today in the dining room.

Table 1.2: Example experimental items in Cunnings and Sturt, 2018

of the sentence containing them, similar to the illusions of grammaticality seen in other studies (i.e., Bock and Miller, 1991, Pearlmutter et al., 1999, Wagers et al., 2009) and the semantic attraction in Laurinavichyute and von der Malsburg (2022).

The main finding reported in the results, was a significant interaction of sentence plausibility and distractor plausibility in total reading times ($p<.01$). Implausible sentences had shorter reading times at the critical (i.e., *shattered*) and spillover region (i.e., *today*) when the distractor word was plausible (i.e., *cup*) than when the distractor word was implausible (i.e., *tie*). In plausible sentences however, no such significant effect was observed with the manipulation of the distractor word (whether it was plausible or implausible) not leading to differing reading times at the critical or the spillover region.

This result was considered compatible with the ACT-R cue-based model and was taken as evidence that retrieval interference influences the formation of dependencies and by extension the semantic interpretation of sentences. It was claimed that distractors are retrieved when the retrieval targets do not fully match the set of retrieval cues at the gap, leading to illusion of plausibility effects, where a plausible but ungrammatical object of the verb leads to facilitation during processing.

The cue-based model could explain such a finding such that as seen in figure 1.12, there is a semantic mismatch between the grammatical *letter* and the features at the retrieval probe by not sharing the quality + *shatterable*. That is also true in this condition for the intervening distractor, as it also has the feature - *shatterable*. However, as seen in figure 1.11, despite the grammatical item *letter* still being a semantically implausible match to *shattered*, the intervening distractor word *cup*, is semantically fitting to it. Therefore it is suggested that processing is facilitated in sentences like the ones in figure 1.11, due to that match.

However, what is not taken into consideration in this study is the structure of the sentences used as experimental items, and whether that plays a role by being the reason for the facilitation behind the observed illusions of plausibility. The sentences used in the study are structured in a potentially ambiguous way. The verb *shattered*

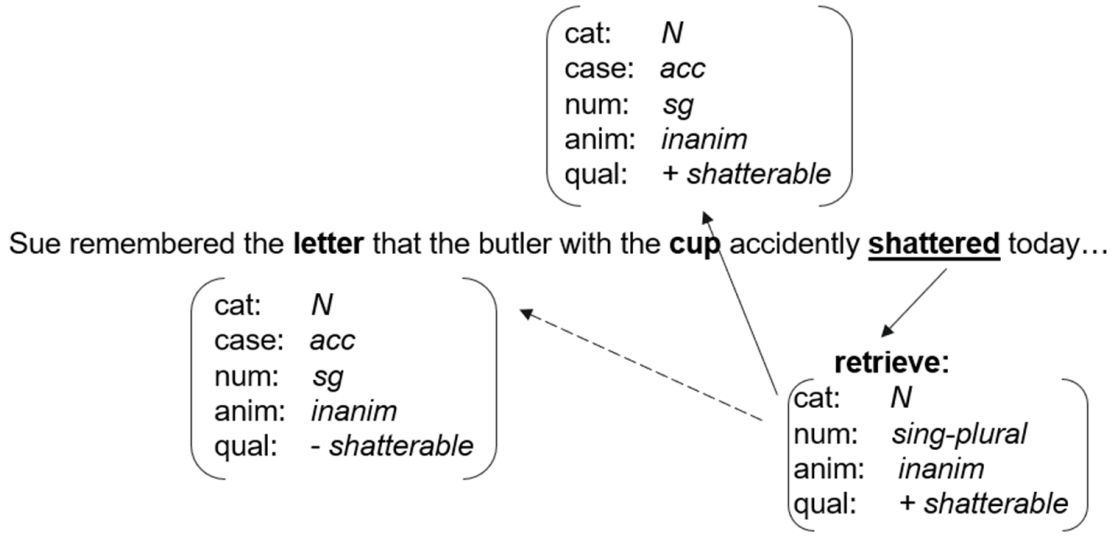


Figure 1.11: *Example of sentence in Cunnings and Sturt (2018) as stored chunk of features in the condition in which illusions of plausibility were found (that is implausible sentence with plausible distractor).*

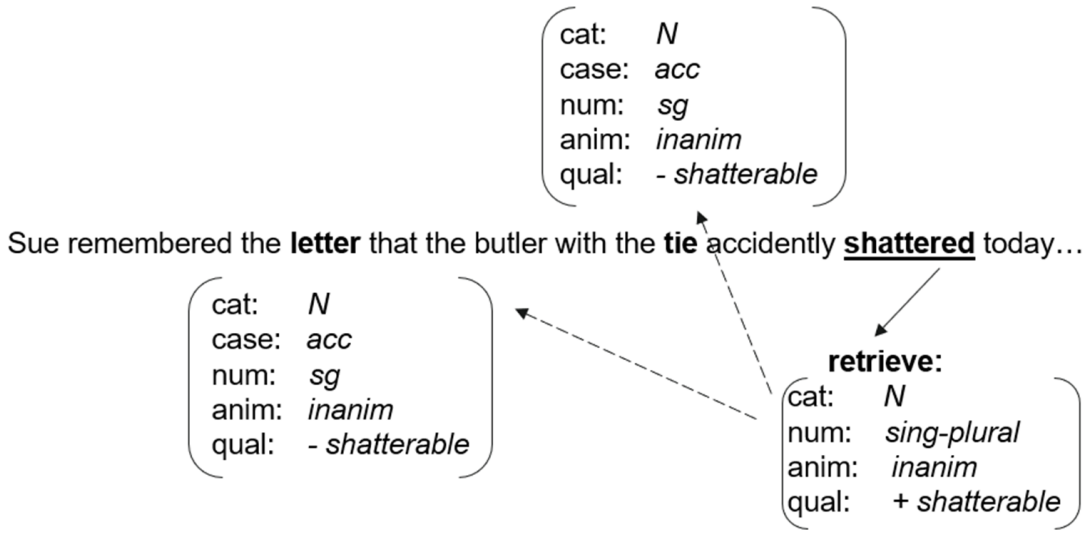


Figure 1.12: *Example of sentence in Cunnings and Sturt (2018) as stored chunk of features in the condition of implausible sentence - implausible distractor.*

in the sentences where an effect was found (example in figure 1.11) could either be interpreted as the main verb of the sentence which fits the explanation in a cue-based framework, or interestingly it could be interpreted as a locally coherent past participle due to its being close to the noun *cup* (i.e., being a reduced version of *the cup (that was) accidently shattered*), which then could not be explained as the reason behind the

illusions of plausibility.

1.6.3 Outline of Present Thesis

The aim of this part of the thesis was to introduce some concepts that are most prevalent in the psycholinguistic literature. In particular, the Introduction gave an overview of how long distance dependencies are parsed, how working memory is essential in those dependencies' parsing, how those memory mechanisms have been implemented in linguistic models and finally how those cue-based models predict interference effects in long distance structures.

In the next parts of the thesis I will try to explore how sentences with long distance dependencies can be interpreted and processed by the parser, and to what extent their processing and interference phenomena that arise in them can be predicted by the already existing and widely supported cue-based retrieval models.

More specifically, chapter 2 looks into semantic attraction phenomena that are considered to be a result of sentence external distractor words. The presented experiments question that claim while asking the question of whether the cue-based retrieval models can truly explain semantic attraction between elements that are not part of the same sentence.

In chapter 3 this work is further extended by looking into semantic attraction phenomena that are considered to arise by sentence internal distractor words. The set of experiments attempts to investigate whether plausibility effects in filler-gap dependencies always lead to semantic attraction effects and whether structure plays a role during their parsing.

Chapter 4 discusses the implications that arose as a result of the experiments run both in chapter 2 and chapter 3, and attempts to tie the findings of the present study with the broader cue-based retrieval literature as well as offering another potential suggestion that could explain those findings through a different approach.

Chapter 2

2.1 Similarity based retrieval interference from sentence external distractor words

As previously discussed in chapter 1, the primary focus of the present research is on the relationship between memory and linguistic knowledge. The two main concepts that will be discussed here are filler-gap dependencies and the predictions that cue-based models make when it comes to the resolution and the processing of this type of dependency. This chapter will attempt to bring those concepts together by discussing them in relation to a particular study (Van Dyke and McElree, 2006), which is one of the most influential of its kind in the field and its results. This chapter will explore a potential confound of this previous study, while also bringing in concepts like relatedness and structural constraints while parsing sentences. More specifically, the following set of experiments will look closer into similarity based retrieval interference as a result of sentence external distractor words interfering with the resolution of filler-gap dependencies due to semantic attraction.

2.1.1 Dependency resolution

Filler-gap dependencies contain elements that are not in their original canonical position, making them challenging structures for the parser as they cannot be immediately integrated into the parse (E. Atkinson et al., 2018; Frazier & Clifton Jr, 1989; Stowe, 1986; Traxler & Pickering, 1996). Since the filler has moved from its underlying thematic position, it is the parser's task to identify the trace where the filler originally moved from and then establish an integration relationship between the filler and the gap in order to successfully process it. Successful processing then requires for the filler to be assigned a thematic interpretation, which is only possible when the filler is linked back to its gap.

Multiple studies have shown that this mechanism behind filler-gap dependency resolution is driven by structural information (i.e. how and where gaps are posited),

whereas other linguistic or extra-linguistic aspects (argument structure, frequency) do not have the same immediate effect on gap filling. Stowe (1986) testing dependency resolution in sentences with subject, object or prepositional phrase extraction, found that the parser posits a gap as early in the structure as it is syntactically possible. Nevertheless, despite the parser's eagerness to reach a full interpretation quickly, there is no attempt to posit a gap in a grammatically unlicensed position. Comprehenders seem to respect island constructions when positing gaps, showing that the parser does not consider all argument positions as equally possible gap sites (Omaki et al., 2015; Sprouse et al., 2012; Stowe, 1986; Wagers & Phillips, 2014). Meanwhile, Staub (2007) and M. J. Pickering and Traxler (2003) testing direct object-verb dependencies with verbs optionally transitive, showed that frequency information is not considered by the parser as it tries to posit a gap at the earliest position possible. Omaki et al. (2015) testing direct object-verb dependency resolution in sentences where the transitivity of the verb was manipulated, found longer first pass reading at the intransitive verbs. That suggests that the parser attempts to build a dependency between the filler noun and the verb without considering the verb's transitivity.

Taken together, these findings suggest that dependency resolution relies primarily (at least during the early stages of processing) on structural constraints. The parser prioritises syntactic information such as on which position is the most likely or the earliest possible for a gap to be realised, whereas bottom-up information on the lexical properties of the gap, such as its transitivity or frequency information, are not considered (Frazier & Clifton Jr, 1989; Osterhout & Swinney, 1993; Staub, 2010; Traxler & Pickering, 1996; Wagers & Phillips, 2014).

2.1.2 Cue-based models

Cue-based models have generally different predictions when it comes to the processing of filler-gap dependencies (Lewis and Vasishth, 2005). Retrieval models propose that the resolution of a dependency is possible due to the retrieval of objects stored in memory, based on their unique sets of cues (McElree, 2000; McElree et al., 2003). The extremely limited working memory capacity to accommodate more than two elements simultaneously (the input currently attended and one extra element), necessitates for the words in a sentence to be stored as 'chunks' or sets of features in memory, containing information associated to that element (syntactic, semantic, lexical) (Gibson, 1998; Lewis et al., 2006; Van Dyke & Johns, 2012; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2011). Based on those models, a filler is integrated when a match between the features at the retrieval probe at the site of the gap and the stored features of any item in memory is achieved. The element that is successfully retrieved from declarative memory, is the one which provides the best match of retrieval cues

with the probe. After this match, the element is brought back in focus in the working memory, it is reprocessed and the filler-gap dependency is formed.

However, as these models employ a mechanism which searches objects in memory in parallel while processing (Lewis and Vasishth, 2005; McElree et al., 2003), in cases where there are multiple items in memory that fully or partially match a set of retrieval cues, similarity based retrieval interference arises (Vasishth et al., 2008), leading to easier processing in structures with multiple items matching a probe, or inhibitory, leading to a delay in processing (Jäger et al., 2017; Lewis et al., 2006).

Interference effects during processing have been investigated in multiple studies (Cunnings & Sturt, 2018; Fedorenko et al., 2006; Gordon et al., 2002, 2004; Van Dyke & McElree, 2011), with a popular position in this literature being that interference could be an effect caused by any item in a sentence, as long as there is a match between the features of the potential retrieval target item and the retrieval probe at the gap (Cunnings & Sturt, 2014; Van Dyke & McElree, 2006; Wagers et al., 2009). Previous findings have been taken to suggest that the cue-based retrieval model allows even for objects outside the current sentential context to be considered as potential fillers in filler-gap sentences, provided that those objects could be possible predicates of the verb at the gap (Fedorenko et al., 2006; Gordon et al., 2002; Van Dyke & McElree, 2006). Gordon et al. (2002) manipulated three words to be either proper names or role descriptions as a memory load which preceded a sentence, to either match or mismatch the nominal filler in a filler-gap dependency. They found lower accuracy and increased reading times for the matched condition than in the unmatched condition. Fedorenko et al. (2006) manipulating the memory load to either have one or three words while testing subject and object relative clauses, found slower reading times for the object relatives when preceded by the three words memory load. However, in both those studies it is difficult to attribute the interference effects to retrieval processes, as the regions they observe them in either precede the retrieval point or contain the whole dependency. Therefore they might be better explained as similarity based interference during the encoding stage.

Despite some of the previous findings being attributed to encoding (i.e., Fedorenko et al., 2006; Gordon et al., 2002, etc.), by suggesting that any item in memory that matches the retrieval probe cues could lead to interference, these models seem to ignore the fundamental role played by structural restrictions in limiting the possible space of search in dependency resolution. An influential study in the literature providing the strongest support to this claim so far is the Van Dyke and McElree (2006) study. It is considered to support a generalised approach, where any element in memory can interfere with dependency formation, independently from structural restrictions.

2.1.3 Sentence external retrieval interference

Van Dyke and McElree (2006) attempted to straightforwardly apply the cue-based model predictions to the processing of filler gap dependencies. More specifically, they investigated whether extra-sentential material can result in proactive retrieval interference effects during the processing of filler-gap dependencies. In a dual task which involved a self-paced reading paradigm and a memorisation-recall task, they tested object-clefted sentences, while manipulating the memorisation of three nouns, and the plausibility of the clefted verb (whether it was a plausible or implausible predicate to the three memorised nouns) (table 2.1).

+/− **memory load:** table/sink/truck
Non-interfering: It was the boat that the guy who lived by the sea **sailed** on two sunny days.
Interfering: It was the boat that the guy who lived by the sea **fixed** in two sunny days

Table 2.1: Example experimental items in Van Dyke and McElree, 2006

The cue-based model prediction is that if the cues of the three memorised nouns are similar with the cues of the filler noun in matching the cues at the retrieval site at the verb, then a cue overload is expected at the retrieval site with semantic interference leading to a slowdown at the critical verb. Indeed a significant interaction of memory load and verb type is reported, with longer reading times for plausible verbs vs. implausible verbs in the load condition (i.e., longer reading times for the sentence containing *fixed* when preceded by the three words *table, sink, truck*), but no effect in the no load condition ($p<.05$). These results are considered compatible with the cue-based retrieval model, and they are taken as evidence that even sentence external nouns can interfere with the formation of intra-sentential dependencies, when those nouns are plausible arguments of the verb at the retrieval probe. It is suggested that the retrieval process is slowed down due to similarity based interference, where the preceding memorised words share the same feature with the syntactic predicate.

However, this effect has not been replicated in other studies. In a subsequent study, Van Dyke et al. (2014) did not replicate the same effect on reading times, while testing participants with varying levels of working memory capacity on the same materials as in Van Dyke and McElree (2006). Mertzen et al. (2020) ran a cross-linguistic study with an eye-tracking paradigm in three languages (English, German, Russian) with similar manipulations of memory load and verb plausibility. The significant interaction of memory load and verb type was again not observed in any of the three languages, while Bayes factor analyses in all three languages showed strong evidence for models which reflect the null hypotheses that the interaction of memory load and verb plausibility is equal to zero.

Apart from the effect not been replicable in these following studies, there is an-

other confound for the reported results in Van Dyke and McElree (2006). The primary focus in that study appears to be the relationship between the memory items and the main verb of the filler gap dependency. The change in that relationship (whether plausible or implausible) is taken as the reason for the observed effect and as support for the application of the cue-based model in processing. However, while the relationship between the three nouns in the memory load and the main verb changes, what also changes is the relationship between the retrieval target noun and the main verb. The relationship between the filler noun *boat* and the main verb *sailed* in the implausible condition, seems to be closer than the relationship between the noun *boat* and the main verb *fixed* in the plausible condition. Therefore an initial theory is that this difference in the relationship between the direct object noun and the verb could be a potential confound for the reported results, or even be the driving force behind the observed effect.

2.2 Pre-tests and Relatedness

To test whether the initial theory of the different levels of relatedness between the filler noun and the main verb was true, two exploratory experiments were run: a cloze task (Experiment 1) and a lexical decision task (Experiment 2). Both of those tasks would allow for a quick check of whether the initial impression about the relationship between the filler noun and the main verbs being meaningfully different is significant.

The reason for deciding to run both of these pre-tests and not settle for only one of these two, stems from the different measures they offer when it comes to relatedness. These two tasks allow for an exploration of relatedness from two different points.

The cloze task was chosen as a pre-test as it could shed some light on the relationship between the filler noun (i.e., *boat*) and the two main verbs (i.e., *sailed, fixed*) used in the sentences in Van Dyke and McElree, 2006 from a probabilistic perspective. It could provide information about the probability of a certain verb appearing after a certain noun in this particular given context of the sentence, while at the same time allowing for the estimate of the surprisal values of each of those verbs (Levy, 2008). The potential significance of the relatedness between noun and verb in a task like this then, could suggest that when a main verb (i.e., *sailed / fixed*) has a high cloze probability of appearing in a certain context or a low surprisal value, processing is facilitated. And moreover that this facilitation is a consequence of the preactivation of the verb by the closely related filler noun (i.e., *boat*) which has preceded it.

On the other hand, the lexical decision task explored the relationship between the filler noun (i.e., *boat*) and the main verbs (i.e., *sailed / fixed*) from a retrieval perspective. The lexical decision tasks allowed for testing whether the filler noun could be primed by one of the main verbs used in the sentences in Van Dyke and McElree, 2006 (more) over the other. If there was a significant difference between the priming effect of one verb over the other when preceding the same noun, it could be suggested that processing is facilitated. And this facilitation could be due to one of the verbs having a closer relationship to the noun than that of the other verb which allows for the noun's easier retrieval.

So to be able to potentially spot this relatedness effect regardless of its actual direction both of these pre-tests were employed when testing the original items from Van Dyke and McElree (2006).

2.3 Experiment 1: Cloze Task

A cloze task (Taylor, 1953) was employed in order to estimate the surprisal values of potential continuations for the sentence frames and in particular for the chosen verbal forms used in the experimental sentences in Van Dyke and McElree (2006).

2.3.1 Participants

60 native British English speakers were recruited for this task (22 of which male with a mean age of 30 years). In their initial screening by the recruitment platform the participants had self-reported to have no literacy or hearing difficulties, no language related disorders, and normal or corrected to normal vision. Participants gave informed consent before their participation and were then compensated after the completion of the experiment (£1.50).

2.3.2 Materials

A truncated version of the experimental sentences from Van Dyke and McElree (2006) were used in this experiment in the form of to be completed fragments. Participants were given a frame of the experimental sentence from the beginning up to just before what in Van Dyke and McElree (2006) was defined as the critical region, which is the region that involves the critical verb. At the critical region there was a gap which the participants were asked to fill. The prepositional phrase in the spillover region of the original sentences was altogether removed for all sentences. Therefore, the word that the participants were asked to add was the final word of the sentence. The main reason for the removal of the spillover region was for the participants not to be further constrained in their choice of words when filling the required gap.

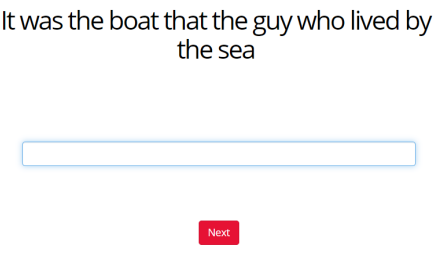


Figure 2.1: Example item in experiment 1 (Cloze task).

2.3.3 Procedure

The cloze task was coded using the Gorilla software (Anwyl-Irvine et al., 2020) and was run via the Prolific Academic platform (www.prolific.co).

The sentence would always appear with a visible gap (___) at the end. The sentential frame would spread in two lines, with the clefted part in one line and the subject with the relative clause in the second line followed by the gap (___). 2 seconds after the appearance of the sentence, a box would appear in the middle of the screen under the sentence where participants could write the continuation they saw most fitting. An example of what the participants saw is given below (2.1). The task was an open completion task, as the subjects were not given any options to choose from, but were directed to complete the incomplete sentence probes in a meaningful way that would be most natural to them. They were explicitly asked to write the word that they thought would fit best as quickly and as accurately as possible. They were not specifically directed to add a verbal form, although the most likely completion for the sentence fragments was indeed a verb, and that was also confirmed at the analysis stage, as the vast majority of the responses included a verb at the critical position.

For the completion of the experiment participants needed approximately 15 minutes.

2.3.4 Analysis

Initially, the responses were inspected by examining how many instances of the two verbs used in the two conditions of the experimental sentences in Van Dyke and McElree (2006) were given in the responses. Responses that included the critical verb in either condition (related or unrelated) but in a different tense than the simple past expected, were taken as an instance of that verb and were included in the analysis (i.e., participants completing the example sentence presented in figure 2.1 with *was sailing* instead of the expected *sailed*).

The percentage out of 100 was estimated based on the instances of each of the expected words in both the related and the unrelated condition (The raw numbers were out of 60, as that was the total number of participants). Then the surprisal value for each one of those was estimated in R (Team R Core, 2020), by calculating the -log probability for each item. To estimate the surprisal value of words that had no tokens in the task, 0 was replaced with the minimum possible value closer to 0 (0.01), assuming that the probability of encountering these words would be extremely low (more specifically 1%).

The surprisal values were then analysed with a Bayesian linear model (Bürkner, 2017a, 2017b; B. Carpenter et al., 2017) that assumes a Gaussian distribution of the

dependent variable, which was the relatedness of the verbs. The default priors were used as no priors were set for this simple analysis. The levels of the factor Relatedness were coded as -1 (related) and 1 (unrelated).

2.3.5 Results

Condition	Cloze Probability	Surprisal Value
Related	.13	2.88
Unrelated	.007	4.41

Table 2.2: *Mean cloze probabilities and surprisal values for related and unrelated verbs in experiment 1 (Cloze task).*

The mean cloze probabilities and the mean surprisal values per condition are given in table (2.2). The mean of the posterior distributions of the fixed effect parameter in this task are summarised in table (5.1, p.138) along with the 95% highest posterior density interval (HPDI). The posterior distribution is given in figure (5.1 p.143). The analysis of the data showed that the posterior distribution of the main effect of Relatedness on the surprisal values of each verb has a range of positive values only ($\hat{\beta}$= 0.76, 95% CrI = [0.50, 1.02]), suggesting that the related and the unrelated verbs were not treated similarly.

2.3.6 Discussion

The model run on the results of the simple cloze task indicates that there is a significant difference in how expected each of the two verbs used in Van Dyke and McElree (2006) in the context of the filler noun. The results indicate that people are more likely to continue a sentence with a verb proposed here to be more related to the filler noun than a verb not so closely related, as manifested in the higher surprisal values of the verbs like *fix*, which is not as closely related to *boat* as *sail* is. That is also confirmed by the plethora of zero instances of the unrelated verbs in the responses in the task, as for many of the unrelated verbs there were no participants choosing to continue a sentence with such a verb. That indicates that it is more likely expecting a related verb to come after a certain noun than an unrelated verb, suggesting that this could potentially affect the results in a given study.

To further explore this relationship, but from a different perspective, a second task was run: experiment 2 (Lexical Decision Task).

2.4 Experiment 2: Lexical Decision Task

The goal of experiment 2 was to further test the relationship between the filler noun and the main verb of the sentences in Van Dyke and McElree (2006), by running a simple lexical decision task.

2.4.1 Participants

60 native British English speakers were recruited via Prolific Academic for this study (25 of which males with a mean age of 28.7 years). The participants self-reported not to have any hearing, vision or literacy difficulties and no language related disorders. They all gave informed consent and were then compensated for their participation in the study (£1.50).

2.4.2 Materials

A total of 72 critical pairs of prime-target words were constructed. The filler nouns from Van Dyke and McElree (2006) were used as target words for 'real word' answers of this priming task. Three lists were created consisting of an equal number of target strings that were non-words and of actual words (36 each). In each list each of the 36 filler nouns (used as target words) was preceded by a prime word which was either one of the two verbs which appeared in the same sentence with the filler noun in the experimental sentences in Van Dyke and McElree (2006) (i.e. SAIL - boat, FIX - boat), or it was a noun (i.e. SYSTEM - boat). The nominal prime words were always unrelated to the target noun and were used as controls, against which the two conditions with the verbal prime words would be compared. The 36 prime - non-word pairs consisted of target words generated by Wuggy (Keuleers and Brysbaert, 2010) and paired with unrelated prime words (i.e. VISIT - hantwase). The non-words generated by Wuggy are strings of letters which resemble real words with a letter or two changed. All the prime word - non-word pairs were repeated across the three lists. The three practice pairings were also repeated across the three lists and consisted of a nominal prime and a target word, a verbal prime and a target word, and a target word and a non-word pairs.

2.4.3 Procedure

The task was coded using the Gorilla software (Anwyl-Irvine et al., 2020) and was run on the Prolific Academic platform (www.prolific.co).

Each trial would begin with a fixation cross in the middle of the screen which remained there for 500ms. The cross would be followed by the prime word in upper-

case letters which was on the screen for 150ms, after which there was a blank screen for 350ms. Then the target string of letters (either actual word or non-word) would appear in lower-case letters and participants were asked to make a judgement on the second word in order to continue to the next trial. If no judgement was made in the first 3000ms after the appearance of the target word, the next trial would begin automatically.

Participants were asked to focus on the fixation cross each time, then read the two strings of letters (prime word + target word/non-word) for each trial and eventually decide whether what they saw in small-case letters was an actual word in English or not, as accurately and as quickly as possible. To register their response they were asked to place their fingers on the letters "j" and "f" on their keyboard and press "j" if the second word was a real word or "f" if it was not. For each one of their judgements in each trial they would see feedback in the form of a green tick if they got it correctly or a red cross in case they made a mistake.

Each session started with three practice trials which were not used in the analysis and did not contain any of the words used in the experiment neither as prime nor target words. All pairs were presented in a random order.

The entire experiment lasted approximately 10 minutes.

2.4.4 Analysis

The collected data were analysed in R (Team R Core, 2020) and a Bayesian linear regression model was fitted to the data using the brms package (Bürkner, 2017b, Bürkner (2017a)). Only trials where a response was given were included in the analysis, while extreme values and outliers were also excluded (2.5 SDs from the by subject mean). The priming effect was estimated for each of the experimental items for each of the conditions, by subtracting the control condition (nominal prime - target noun, i.e. SYSTEM - boat) from each of the experimental conditions (verbal prime - target noun, i.e. FIX - boat, and SAIL - boat). A Bayesian linear model which assumes a Gaussian distribution of the dependent variables was fitted to the data to test for priming effects. The two variables in the model were the two conditions of interest, namely the related (i.e. SAIL - boat) and the unrelated (i.e. FIX - boat). The levels of each were coded as 1 (if belonging in the category) and 0 (if not).

2.4.5 Results

The mean priming effects per condition and the mean reaction times are given in table (2.3) The standard error is given in parenthesis. The mean priming effect per condition is plotted in figure 2.2. The mean of the posterior distributions of the parameters

Condition	Reaction Time (ms)	Priming Effect (ms)
Related	641.03 (46.72)	-39.77
Unrelated	665.70 (42.29)	-33.56

Table 2.3: *Mean reaction times (ms) and priming effects (ms) in the related and the unrelated conditions in experiment 2 (Lexical Decision Task). The SE is given in parenthesis.*

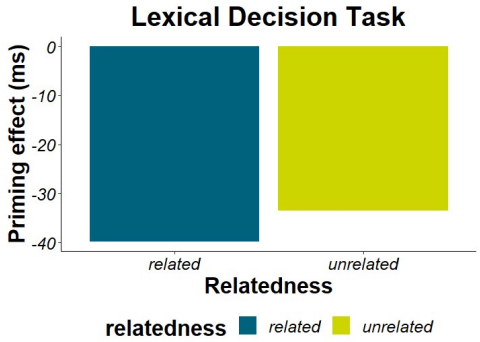


Figure 2.2: By condition mean priming effect in experiment 2 (Lexical decision task).

are given in table (5.2, p.138), as is the 95% highest posterior density interval (HPDI). The plotted posteriors are given in figure (5.2, p.143). The posterior distribution for the effect of Related verb type showed a credible interval with negative values only ($\hat{\beta}$= -55.18, 95% CrI = [-100.72, -12.26], P($\hat{\beta}$ < 0) = 0.99), while that for the effect of Unrelated verb type showed a credible interval like of ($\hat{\beta}$= -30.71, 95% CrI = [-75.05, 13.13], P($\hat{\beta}$ < 0) = 0.91) suggesting overall shorter reaction times for the noun when preceded by a related verb.

2.4.6 Discussion

The analysed data of the lexical decision task are consistent with the data of the previous cloze task. The analysis suggests that when a noun (one of the ones used as filler nouns in Van Dyke and McElree (2006)) is preceded by a verb more closely related to it (i.e. *boat* preceded by *SAIL*), then there is facilitation in the reaction times in participants' responses. This effect then is different from the one found for the verbs not as closely related (i.e. *boat* preceded by *FIX*), where though different from the control numerically, the effect is not significant based on the statistical analysis model.

2.5 Thematic Integration Account

A significant effect of relatedness was found in both of the pre-tests. The semantically related verbs (i.e. boat - **sail**) had significantly lower surprisal values than the semantically unrelated verbs (i.e. boat - **fix**) when following the filler nouns, while the semantically related verbs were also found to prime the filler nouns (i.e. SAIL - boat) more than the unrelated verbs(i.e. FIX - boat). The cloze task and the lexical decision task supported the theory set out earlier that there is a different semantic relationship between the retrieval target filler (boat) and the two retrieval verbs (sail vs. fix), which should not be overlooked when interpreting the data in the previously discussed studies.

Based on the results of those two exploratory studies then, a question arises of whether the results from Van Dyke and McElree (2006) are indeed due to cue based retrieval interference caused by sentence external distractors or whether there is an alternative explanation. What is proposed here is an alternative account which could offer an explanation for their findings, but which is totally independently from cue-based retrieval: the thematic integration account. The thematic integration account focuses on the relationship between the elements involved in the dependency, suggesting that a close semantic relationship between a filler noun and a verb can facilitate thematic integration.

This facilitation during integration could be have two possible explanations. One possibility could be that this facilitation is a retrieval advantage of words which are closely semantically related, which would suggest that the antecedent 'boat' is retrieved more easily, and thus faster, when retrieval is triggered by the more closely related verb 'sail' in comparison to when it is cued by the verb 'fix'. Alternatively, it could be an effect of preactivation, as in the verb 'sailed' at the gap being preactivated by the filler noun 'boat' in a larger degree than the verb 'fixed', which consequently leads to easier and faster integration.

From the previous two experiments in this study it is not clear what relatedness encompasses, whether it is about the preactivation of the verb or the easier retrieval of the filler noun. However establishing which one of the two views can better explain the relatedness effect is not crucial neither for the thematic integration account which is proposed here, nor for the experiments to follow. In any case what is and will be discussed as relatedness for the remainder of this study is the closer relationship between the filler noun and the main verb regardless of the direction that this effect might have, whether it signals preactivation or retrieval.

The main result in Van Dyke and McElree (2006) was the significant difference between the processing of the implausible over the plausible condition in the memory load condition only (i.e., so longer reading times for the verb *fixed* when preceded by

the memory load words *table, sink, truck*, compared to the verb *sailed* when preceded by the same memory load words), while there was no verb type effect in the no load condition. Initially, this result does not seem to be fully compatible with the proposed thematic integration account, as the account predicts an effect to be found when the filler and the retrieval verb are semantically related (i.e., *boat, sailed*), suggesting that the facilitation effect should be found in both load and no load conditions. However, the lack of an effect in the no memory load condition, could suggest that the discussed effect is more easily detectable when there are increased task demands (i.e. dual task involving memorising a list of words in a self-paced reading design). The increased complexity in the memory load condition could have accentuated the contrast between the verb type conditions, leading to an effect in the load condition only. If that is true, it would mean that the memory list is not causing similarity-based interference, but is instead inducing a task-based effect.

Overall then, the previous results could be explained by the thematic integration account, considering 1. that in the thematic integration account priority is given to the structural information in the sentence (filler noun - verb) and 2. if potentially such an effect might be a direct result of the change in the relationship between items in structurally licensed positions in the sentence, which leads to significant effects in taxing enough tasks.

To explore those further and to test whether the thematic integration account could explain the resolution of filler gap dependencies, three experiments were run: a dual task self-paced reading task (Experiment 3), a maze task (Experiment 4), and a simple self-paced reading task (Experiment 5).

2.6 Experiment 3: Dual Task Self-paced Reading

To test whether sentence external distractors can lead to interference effects in filler gap dependency resolution or if resolution is heavily based on structural information, a dual task self-paced reading study was run. The thematic integration account predicts that running sentences with varying levels of relatedness in a taxing parsing environment would allow for the relatedness effect to be observed. Holding the presence of a memory load constant in all conditions allowed for the increased complexity of the task.

If the results of Van Dyke and McElree (2006) and Van Dyke et al. (2014) were due to similarity based interference, an interaction between the original verb manipulation and load type should be expected in the current experiment, such that longer reading times at the critical region of the verb (online), and/or slower response times and lower accuracy in comprehension questions (offline) for sentences in the unrelated with nominal load condition. At the same time no difference should be expected in the conditions with the verbal load.

In contrast, if the previous findings were due to facilitated integration for semantically related filler noun-verb pairs compared to pairs that are unrelated, an effect of relatedness would be expected across the load type manipulation. That would mean shorter reading times at the verb, shorter response times and/or higher accuracy in comprehension questions should be expected in sentences in the related condition independently of the words in the memory load list.

2.6.1 Participants

60 first year University of York students were recruited for this study. All participants were native British English speakers, had normal or corrected to normal vision and were given course credit for one of their modules for their participation. All participants gave written consent before their participation in the study.

2.6.2 Materials

The original stimuli, fillers and comprehension questions from Van Dyke and McElree, 2006 were used in a 2x2 factorial design crossing memory load type (nominal vs. verbal) and relatedness (related vs. unrelated).

There were two main differences between the materials in this and the materials in the original study (Van Dyke and McElree, 2006). The first difference is in the main verb manipulation. In this experiment the main verb was characterised based on its relationship with the filler noun. The experimental sentences had two conditions characterised as *related* and *unrelated*. The sentences in the related condition

contained filler nouns and main verbs with a closer semantic relationship (i.e., *boat, sailed*), as opposed to sentences in the unrelated condition, where the filler noun and the main verb are not as closely related (i.e., *boat, fixed*) as seen in table 2.4. That is unlike the original study (Van Dyke and McElree, 2006), where the main verb manipulation was characterised on the basis of the relationship between the verb and the contents of the memory load. Namely in the interfering condition for sentences containing the verb *fixed*, the contents of the memory load (*table, sink, truck*) are assumed to potentially interfere with the resolution of the dependency by being possible good matches for the verb *fixed*, as they are *fixable* (table 2.1). In the non-interfering condition however, in sentences containing the verb *sailed*, the contents of the memory load (*table, sink, truck*) cannot be considered as potential matches for the verb *sailed* by not being *sailable*, and therefore were not expected to interfere with the resolution of the dependency in the sentence.

In particular, a sentence like: *It was the boat that the guy who lived by the sea* ***sailed*** *in two sunny days.* preceded by a memory load of *table/sink/truck*, is a sentence that in Van Dyke and McElree, 2006 was in the condition named non-interfering as the memory load items (*table/sink/truck*) cannot interfere with the verb *sailed*, whereas in this study the same sentence is in the related condition as the verb *sailed* and the filler item *boat* in it are closely semantically related. Similarly, a sentence like: *It was the boat that the guy who lived by the sea* ***fixed*** *in two sunny days.* preceded by a memory load of *table/sink/truck* was in the condition named as interfering in Van Dyke and McElree, 2006 with the memory load items potentially interfering with the resolution of the dependency containing *fixed*, whereas in this study the same sentence is in the condition named unrelated due to the verb *fixed* and the filler item *boat* not being as closely semantically related.

The second difference is in the memory load manipulation. The memory load content was manipulated so that it was either nominal as in the original study, or verbal. The words for the nominal memory load were identical to the ones used in (Van Dyke and McElree, 2006). The words for the verbal memory load were all newly generated. A similarity value between each verb in the memory load and the main verb in the sentence was calculated for each experimental sentence. Then an average similarity rating was calculated for both conditions (the related condition and the unrelated condition), whilst ensuring that those were not statistically different. The average similarity was calculated using the WordNet::Similarity database (Pedersen et al., 2008), which is based on the original WordNet corpus (Fellbaum, 2010; Miller, 1995). This database allows for the computation of the similarity between two different concepts (Pedersen et al., 2004) by using gloss vectors. These gloss vectors combine content from the dictionary as well as corpus data in order to generate a similarity score between two concepts. This measure was used as it correlates strongly

with human judgements of semantic relatedness (Patwardhan, 2006).

In the related condition the three words of the nominal load were always implausible arguments of the main verb, whereas in the unrelated condition all three words in the nominal load could be plausible arguments of the verb. The contents of the verbal memory load were always implausible arguments of the main verb, regardless of the condition of the main verb (related or unrelated). An example experimental item is given in table (2.4).

As in the original study (Van Dyke and McElree, 2006), comprehension question responses were split evenly between positive ("Yes") and negative ("No") answers.

memory load: ***nominal:*** table/sink/truck or ***verbal:***chop/learn/try
Related: It was the boat that the guy who lived by the sea **sailed** in two sunny days.
Unrelated: It was the boat that the guy who lived by the sea **fixed** in two sunny days

Table 2.4: Example of experimental item in dual task self-paced reading experiment

All the experimental stimuli for the experiment can be found on p.152.

2.6.3 Procedure

The experiment was coded using the Experiment Builder software (Builder, 2007; Research, 2004). Four counterbalanced lists were created so that each item would appear in one condition in each list. Each trial began with the presentation of the three words of the memory load in one line, separated by dashes. Subjects were asked to read the three words out loud and try to memorise them. Participants then had to read a sentence phrase by phrase in a self-paced reading design. The sentence was then followed by a comprehension question to test the participants' understanding of the sentence. In order to respond, subjects had to press the 'f' key if the correct answer was on the left side of the screen, or the 'j' key if it was on the right side. After responding to the comprehension question, a screen with three rectangular boxes would appear, prompting participants to recall and write down the three memorised words in the order they had originally appeared. If they could not remember them in order, they were asked to write them down in any order they chose to. A new trial would begin after participants filled the boxes with the three words they had memorised or after pressing the key to continue in case they did not remember all or some of the words. The experiment lasted approximately 60 minutes.

2.6.4 Analysis

Reaction times of below 100 ms or above 10.000 ms were excluded from the data analysis. A minimum threshold of 75% accuracy on comprehension questions was also set, leading to the exclusion of data from 3 participants.

For the analyses, Bayesian linear mixed effects regression models were fitted to the data using the brms package (Bürkner, 2017a, 2017b) in R (Team R Core, 2020), implemented in the programming language Stan for Bayesian estimation of model parameters (B. Carpenter et al., 2017, Stan Development Team, 2020). Response data and reading times were log-transformed and analysed with a Bayesian linear mixed effects model that assumes a Gaussian distribution of the dependent variables, whereas a Bayesian linear mixed effects model with a Bernoulli link function was fitted to the accuracy data.

The models included the main effects of Load Type, Verb Type and their interaction (Load Type x Verb Type) as fixed effects. Participants and items were treated as random effects for each of the models. The random effects part of the models included adjustments for subjects and items of an intercept, of slopes for the effects and interaction of verb and load types and the correlations between intercepts and slopes. The levels of the factor Verb Type were coded as -1 (related), and 1 (unrelated) and the factor Load Type was coded as 1 (nominal) and -1 (verbal).

To avoid overfitting or placing strong constraints on the models' predictions, mildly informative priors with a regularizing effect on the posterior distributions were chosen. Normal priors were set for all the fixed effects, with a mean value of 0 and a distribution variance set to 1. A standard normal distribution was also used for the intercept, with the mean value set at 0 and the variance set at 10, whereas for the random effects standard deviations (sd) and the residual variance (sigma) a truncated normal distribution with a mean of 0 and a standard deviation of 1 was used. [1] As the random effect correlation between intercept and slope parameter, the LKJ-correlation prior was used. The LKJ-correlation prior sets the parameter value to 2 and has a regularising effect of assuming that extreme values are highly unlikely.

Four Monte Carlo Markov chains with 4000 iterations each were run in parallel for each model. The first 2000 iterations were discarded as part of the warm up period of the model, leaving in total 8000 iterations for each model. To calculate the Bayes factor, four chains were run with each of them having 10.000 iterations, 2000 of which were part of the warm up phase.

To assess the convergence of the models, it was verified that the Rs (the between to within chain variance) were always at 1.0 for all fixed effects parameters, while the chains were also visually inspected.

The 95% quantile based Bayesian credible intervals and mean estimates are reported here. The 95% credible intervals would mark the range of values within which there is 95% certainty that the true value of the parameter of interest falls, given the

[1] For setting priors that were appropriate for the models reported for the experiments that were run for this thesis, advice was taken from the Nicenboim and Vasishth, 2016 paper. How results are reported is also based on the same paper.

model and data.

2.6.5 Regions of Interest

For the purposes of this task, the regions that were primarily analysed and will be reported are two: the critical and the spillover. As the design of this task is only minimally different from the one in the original study (Van Dyke and McElree, 2006), the same regions will be identified as the critical and the spillover regions. That means that the critical region comprises of the main critical verb (i.e., *sailed / fixed*), whereas the spillover region includes the prepositional phrase following the main verb until the end of the sentence (i.e., *in two sunny days.*)

2.6.6 Results

Four models will be reported: two models for response times and accuracy to comprehension questions, as well as two models for online reading times at the two regions of interest: the critical region (verb: i.e. *fixed / sailed*) and the spillover region (phrase following the verb i.e. *in two sunny days*).

Reading Times

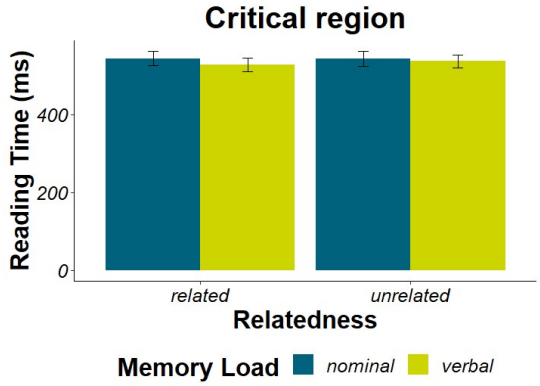


Figure 2.3: *Mean reading times at the critical region in experiment 3. Error bars represent the standard error of the means.*

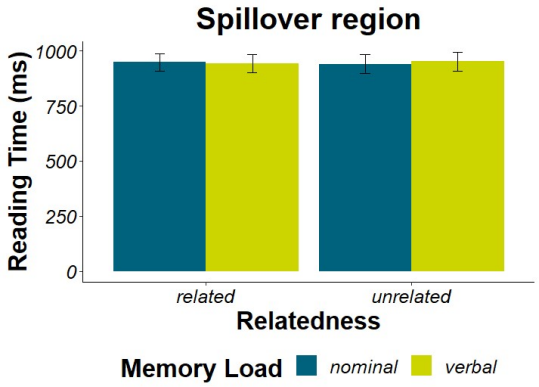


Figure 2.4: *Mean reading times at the spillover region in experiment 3. Error bars represent the standard error of the means.*

Average raw reading times by condition in the critical and spillover regions are given in table (2.5) and plotted in figures (2.3 & 2.4). The standard error is given in the parentheses. The mean of the posterior distributions over the experimental fixed effect parameters in the critical and the spillover regions, along with the 95% highest posterior density interval (HPDI) are summarised in table (5.3, p.138). These Bayesian credible intervals indicate where the most plausible parameter values for these fixed

Condition	Critical region	Spillover region
Related/Nominal	543.96 (18.31)	947.97 (39.60)
Related/Verbal	527.63 (17.20)	941.27 (40.76)
Unrelated/Nominal	543.42 (18.64)	940.32 (43.55)
Unrelated/Verbal	536.84 (17.09)	950.82 (41.99)

Table 2.5: *Mean reading times (ms) per condition by region in experiment 3. The SE is given in parenthesis.*

effects parameters lie, given the data. The posterior distributions are plotted in figure (5.5, p.144) for the critical region and in figure (5.6, p.144) for the spillover region.

In the analysis of the reading times at the critical region, the posterior distributions showed that the estimates of the main effects and the interaction were uninformative, as the credible intervals spanned a range of values below and above zero. Specifically, the Load Type estimate was ($\hat{\beta}$= -0.01, 95% CrI = [-0.02, 0.01], P($\hat{\beta}$ < 0) = 0.83), the Verb Type ($\hat{\beta}$= 0.00, 95% CrI = [-0.02, 0.02], P($\hat{\beta}$ > 0) = 0.53) and the Load Type x Verb Type interaction ($\hat{\beta}$= 0.01, 95% CrI = [-0.01, 0.02], P($\hat{\beta}$ > 0) = 0.82).

The posterior distributions for the main effects in the spillover region were similar providing no evidence for any of the effects, as the credible intervals spanned through zero: Load Type ($\hat{\beta}$= 0.00, 95% CrI = [-0.02, 0.02], P($\hat{\beta}$ > 0) = 0.51), Verb Type($\hat{\beta}$= -0.00, 95% CrI = [-0.02, 0.02], P($\hat{\beta}$ < 0) = 0.52), Load Type x Verb Type ($\hat{\beta}$= 0.01, 95% CrI = [-0.01, 0.03], P($\hat{\beta}$ > 0) = 0.73).

To account for the conflict between the lack of an online effect in the critical region in this task and the reported effect in the critical region in Van Dyke and McElree (2006), further investigation was required. A Bayes factor analysis was computed as it allows for the identification of the model which best describes the effect (or the lack of an effect in this case) in the given data set (Jeffreys, 1998; O'Hagan & Forster, 2004). The present data set was run through two models : the first model assumed a Load Type x Verb Type interaction effect (H_1) whereas the second model assumed there is no difference between the conditions as a result of the interaction of Load Type x Verb Type (H_0). Both of the models included the fixed effects of Verb Type and Load Type but model 1 also included their interaction, whereas the null model did not. All the other parameters were as described above. The Bayes factor analysis indicated that there is very strong evidence against hypothesis 1, $BF_{(10)}$ = 0.0 (the Bayes factor is reported with the model one in the numerator). Base on the Bayes Factor there is 0% likelihood that H_1 is better supported by the data over H_0.

Comprehension Question Accuracy

The mean percentages of correct responses in each condition are illustrated in figure (2.5). Mean response times of correctly answered comprehension questions are shown in figure (2.6). Table (5.4, p.139) summarises the mean of the posterior

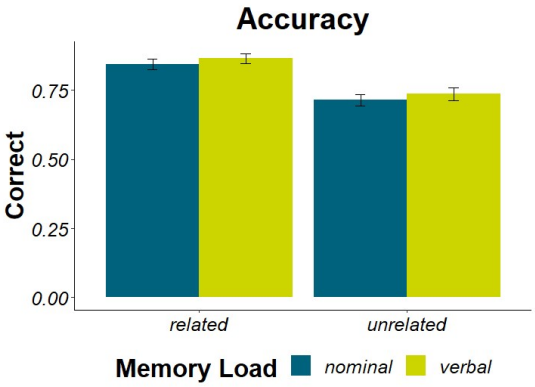


Figure 2.5: *Proportion of correct responses by condition in experiment 3. Error bars represent standard error.*

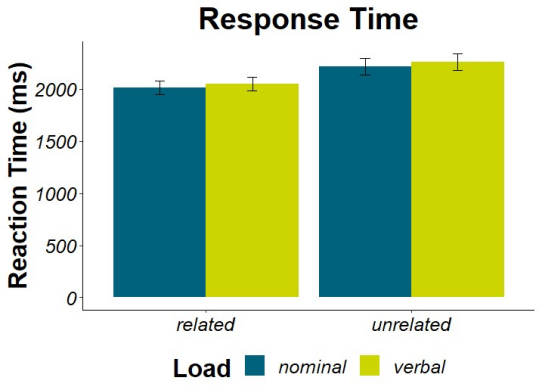


Figure 2.6: *Mean response times (ms) to comprehension questions in experiment 3. Error bars represent standard error.*

distribution over the experimental fixed effect parameters and the 95% CI in the accuracy model and table (5.5, p.139) summarises the same for the response times model. The posterior distributions are plotted in figure (5.3, p.143) for the accuracy model and in figure (5.4, p.143) for the response times model.

The analysis of the comprehension question accuracy data showed that the posterior distribution for the main effect of Verb Type has a range of negative values only ($\hat{\beta}$= -0.50, 95% CrI = [-0.93, -0.09], P($\hat{\beta}$ < 0) = 0.99), suggesting that subjects were more accurate in their answers in comprehension questions following sentences with a verb in the related condition than sentences with a verb in the unrelated condition. There was no evidence for the main effect of Load Type ($\hat{\beta}$= -0.13, 95% CrI = [-0.31, 0.05], P($\hat{\beta}$ < 0) = 0.92) and no evidence of an interaction between Load Type x Verb Type ($\hat{\beta}$= 0.04, 95% CrI = [-0.14, 0.22], P($\hat{\beta}$ > 0) = 0.66).

Comprehension Question Response Times

The analysis of the comprehension question response times data followed a similar pattern. The posterior distribution for the main effect of Verb Type has a range of positive values only ($\hat{\beta}$= 0.04, 95% CrI = [0.02, 0.06], P($\hat{\beta}$ > 0) = 0.99), suggesting that subjects were faster in their responses to comprehension questions following sentences with a verb in the related condition than sentences with a verb in the unrelated condition. As predicted, there was no evidence for the main effect of Load Type ($\hat{\beta}$= 0.01, 95% CrI = [-0.00, 0.02], P($\hat{\beta}$ > 0) = 0.92) and no evidence of an interaction between Load Type x Verb Type ($\hat{\beta}$= 0.00, 95% CrI = [-0.01, 0.01], P($\hat{\beta}$ > 0) = 0.52).

2.6.7 Discussion

The analysed results of this task were partially compatible with the expectations set out before running it.

Overall the offline results in this task were compatible with the predictions of the thematic integration account. The strong relatedness effect both in comprehension question accuracy (manifested as higher accuracy for the questions following sentences in the related condition independently from the memory load manipulation), as well as in comprehension question response times (manifested as faster response times for the related condition) were in agreement with what was expected by the proposed account. At the same time these results were also consistent with the results in Van Dyke and McElree (2006) and Van Dyke et al. (2014), and they would not necessarily be against the expectations set by the cue-based retrieval models. A possible explanation of these results in a cue-based framework, could be that generally the sharing of features should facilitate processing. In the related condition the verb and the filler noun would be expected to share more or actually differently weighted cues, including some type of semantic cue that is weighted more than the cue shared by words that are not as closely semantically related. Essentially that would mean that in the related condition, the verb *sailed* and the filler noun *boat* probably share a semantic proximity feature that leads to facilitation in retrieval (at least to a larger extent than the sharing of semantic proximity cue between the verb *fixed* and the noun *boat* in the unrelated condition) which could be what was manifested in the offline measures that were reported in the previously discussed experiments.

However the online effect –or the lack of any effect to be more precise– was not compatible with neither of the two accounts discussed. The effect found in Van Dyke and McElree (2006) was not replicated in this task and the Bayes factor analysis confirmed that such an interaction could not be true for such a data set given the priors of the model that were set for the analysis. Therefore there was no evidence of interference between the memory load manipulation and the relatedness of the main verb, manifested in faster reaction times for the critical region in the unrelated with a nominal memory load condition, which is against what the cue-based models would predict. At the same time though, the lack of a relatedness effect was not compatible with the predictions made by the thematic integration account, which predicted that the reaction time for the verb at the critical region in the related condition should be faster than that of the verb in the unrelated condition. Since no effect reached significance in any of the models in the critical or even in the spillover region, the question remaining was whether the effect predicted by the thematic integration account could be found online.

2.7 Experiment 4: Maze Task I

As there were no online effects in the Dual Task Self-paced Reading experiment, the next task was an attempt to find online effects of relatedness with a more sensitive and temporally precise paradigm, the maze task (Boyce et al., 2020; Forster et al., 2009; Witzel et al., 2012). The memory load manipulation which was present in the previous dual task self -paced reading study and in Van Dyke and McElree (2006) and Van Dyke et al. (2014) was not deemed essential to include here, as the difficulty added by that manipulation was already provided by the maze task itself.

The maze task is a relatively unnatural design which requires participants to make choices while building sentences at their own pace. In the maze task, subjects are progressing through a sentence by choosing which word of the two they are presented with provides a better continuation for the current sentence in each position. Choosing the wrong word at any stage immediately terminates the sentence, and a new trial begins. In order to progress from one position in the sentence to the next, participants are forced to process the sentence string to make informed and accurate choices for each position. Therefore, the memory load manipulation was removed 1. to make the task more feasible and 2. because its presence is not essential for testing integration facilitation between the clefted verb and a semantically related object argument, as the design itself ensures incremental parsing.

As the maze task is a more precise method than the self-paced reading task which was used before, an effect of relatedness was expected to be found online manifested in faster reaction times in the critical region for the verbs in the related condition, in line with what the thematic integration account predicts. The offline results were also expected to replicate the offline effects in the previous dual task self-paced reading, as in higher accuracy and faster response times for the sentences in the related condition.

2.7.1 Participants

48 native British English speakers were recruited to complete the experiment via the Prolific Academic platform. All participants had self-reported to have normal or corrected to normal vision, no hearing or literacy difficulties and no language related disorders. Before their participation subjects gave informed consent and after the end of the task they were compensated for their participation (£5.63).

2.7.2 Materials

The same stimuli, fillers and comprehension questions from Experiment 3 were used in this task. The same main verb relatedness manipulation was used, with the conditions being characterised based on the relationship between the filler noun (i.e., *boat*)

and the main verb (i.e., *sailed / fixed*). In the related condition their semantic relationship was closer (i.e., *boat, sailed*), whereas in the unrelated condition it was not (i.e., *boat, fixed*). The manipulation of the memory load was altogether removed.

The maze design required for an alternative word (foil word) to be presented alongside the target word for each position of each experimental and each filler sentence. The A-maze software was used to generate the foils for all the words in the experimental and the filler items (Boyce et al., 2020). After being generated by the A-maze software, the foils were manually checked and occasionally altered to avoid having two plausible continuations for a position in the sentence, with a particular focus at the region of interest and the spillover region. The foil words were kept constant for each position for both conditions of each experimental sentence.

Relatedness	**Sentence**
related	It was the boat that the guy who lived by the sea **sailed** in two sunny days.
unrelated	It was the boat that the guy who lived by the sea **fixed** in two sunny days.

Table 2.6: Example of experimental item in maze task

All the experimental stimuli for the experiment can be found on p.160.

2.7.3 Procedure

The experiment was coded using the Ibex Farm software (Zehr and Schwarz, 2018b, Drummond, 2013) and was run via the Academic Prolific platform. In order to encourage high levels of attention to the task, participants could only use a desktop device to run the experiment. Two counterbalanced lists were created, so that each item appeared in only one condition in each list. Participants were required to read sentences one word at a time. Each trial began with the first word of the sentence appearing on the left side of the screen, whereas on the right side there were three "x"s separated by dashes (x-x-x). For each subsequent word for each position, participants were required to make a choice between two words appearing on the left and on the right side of the screen, with one of those words being a better continuation of the sentence than the other. The target word was always allocated a side randomly during the trials for each condition and each participant minimising the risk of an effect due to the target word's position. To make a choice between the target and the foil words, participants were asked to press the 'e' key on their keyboard for the word on the left or the 'i' key for the word on the right. Choosing the wrong continuation word at any point for any of the positions in the sentence would terminate the currently built sentence, and a new sentence would appear. Each sentence was then followed by a comprehension question, to test participants' understanding and attention to the task. To answer the comprehension question participants had to press

the 'e' key on their keyboard for the answer on the left or the 'i' key for the answer on the right. A new trial would begin after answering the comprehension question. The experiment lasted approximately 45 minutes.

2.7.4 Analysis

Prior to the analysis a threshold of 75% accuracy in comprehension questions was set. One participant did not meet this criterion and was therefore not included in the analysis. To control for extreme reaction times, any reaction time above 10.000 ms was excluded from the data analysis. Trials where the incorrect foil word was selected in any position prior to the critical word position were excluded from the analysis of the online data, whereas all trials with incorrect word selections in any positions were excluded from the analysis of the offline data.

The levels of the factor relatedness were coded as -1 (related), and 1 (unrelated). A high probability that the relatedness parameter was smaller than zero would suggest that there was evidence that the related condition was processed faster and / or more accurately than the unrelated condition. By contrast, a high probability that the parameter Relatedness is greater than zero would mean that the condition where the verb is unrelated was processed more accurately / faster.

All other statistical analysis procedures and parameters were identical to those of experiment 3.

2.7.5 Regions of Interest

Two regions were analysed and will be reported: the critical and the spillover regions. The critical region involved the choice that participants made for the main verb of the sentence (i.e. *sailed / fixed*) which was consistent with the the critical region in experiment 3. However, the spillover region was not the prepositional phrase as in experiment 3 but just the one word in the position right after the main verb (i.e. *in*).

2.7.6 Results

The four models reported are on the response times and accuracy to comprehension questions, as well as two models for online reading times at the two regions of interest: the critical region (verb: i.e. *fixed/sailed*) and the spillover region (word following the verb: i.e. *in*).

Reading Times

Average raw reading times by condition in the critical and spillover regions are given in table (2.9) and plotted in figures (2.7 & 2.8) Table (5.6, p.139) summarises the

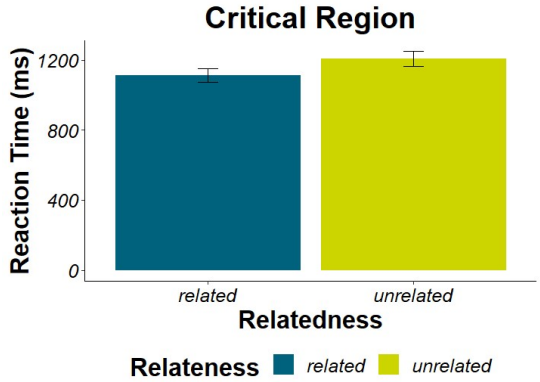


Figure 2.7: *Mean reading times at the critical region in experiment 4. Error bars represent the standard error of the means.*

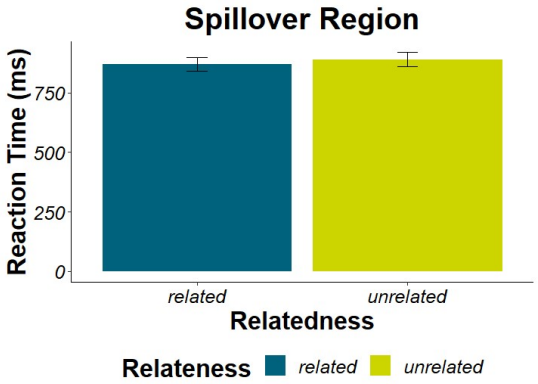


Figure 2.8: *Mean reading times at the spillover region in experiment 4. Error bars represent the standard error of the means.*

Condition	Critical region	Spillover region
Related	1111.39 (40.12)	870.59 (29.22)
Unrelated	1205.82 (43.47)	890.32 (30.50)

Table 2.7: *Mean reading times (ms) per condition by region in experiment 4. The SE is given in parenthesis.*

mean of the posterior distribution over the fixed effect parameter in the critical and the spillover regions, along with the 95% highest posterior density interval (where the most plausible parameter values for the Relatedness effect lie). The posterior distributions are plotted in figure (5.9, p.145) for the critical region and in figure (5.10, p.145) for the spillover region.

As predicted, the model for the reading times in the critical region provided evidence for a change in reading patterns caused by the experimental condition. The posterior distribution for the effect of Relatedness showed a credible interval with positive values only ($\hat{\beta}$= 0.09, 95% CrI = [0.03, 0.15], P($\hat{\beta}$ > 0) = 0.99), suggesting longer reading times at the critical region for the unrelated verb as opposed to the related verb.

The posterior distribution for the reading times in the spillover region showed no effect of Relatedness ($\hat{\beta}$= 0.03, 95% CrI = [-0.01, 0.07], P($\hat{\beta}$ > 0) = 0.92).

Comprehension Question Accuracy

Figure (2.9) summarises the mean percentages of correct responses in each condition. Mean response times of correctly answered comprehension questions are shown in figure (2.10). The mean of the posterior distribution over the Relatedness effect and the 95% CI in the accuracy model is summarised in table (5.7, p.139) and table (5.8, p.139) summarises the same for the response times model. The posterior distributions are plotted in figure (5.7, p.144) for the accuracy model and in figure (5.8, p.144)

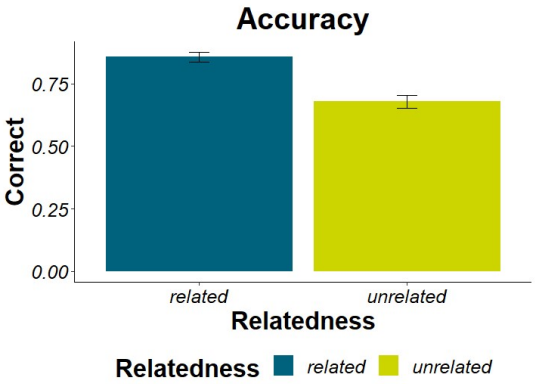


Figure 2.9: *Proportion of correct responses by condition in experiment 4. Error bars represent standard error.*

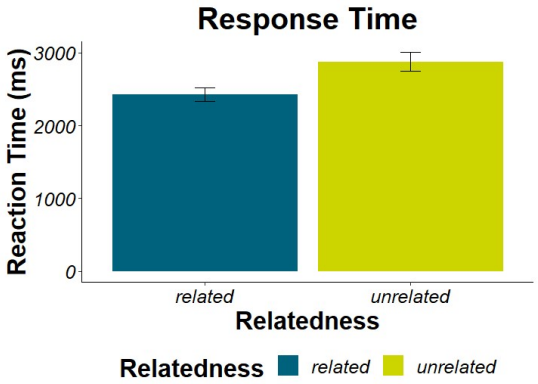


Figure 2.10: *Mean response times (ms) to comprehension questions in experiment 4. Error bars represent standard error.*

for the response times model.

The posterior distribution for the main effect of Relatedness provided evidence for a very strong Relatedness effect. It has a range of negative values only and a very large mean ($\hat{\beta}$= -0.86, 95% CrI = [-1.31, -0.42], P($\hat{\beta} < 0$) = 0.99), suggesting that subjects were more accurate responding to comprehension questions following sentences with a verb in the related condition than to questions following sentences with the verb in the unrelated condition.

Comprehension Question Response Times

The analysis of the comprehension question response times data showed that the posterior distribution for the main effect of Relatedness has a range of positive values only ($\hat{\beta}$= 0.08, 95% CrI = [0.01, 0.15], P($\hat{\beta} > 0$) = 0.98), suggesting that subjects were faster in their responses to comprehension questions following sentences with a verb in the related condition than to questions following sentences with a verb in the unrelated condition.

2.7.7 Discussion

The analysis of the data set of this task is consistent with the predictions made by the thematic integration account.

The attempt to find the effect online was successful, as the maze task allowed for the relatedness effect to be found. As confirmed by the models run, the relatedness effect was particularly strong in the critical region of the sentence, with people being faster selecting the related verb (i.e., *sailed*) over the unrelated verb (i.e., *fixed*) for that region.

The offline relatedness effect found in experiment 3 is replicated with this task as well. Participants were more accurate and faster in their responses to comprehension

questions after the sentence in the related condition, which is compatible with the thematic integration account's prediction set out originally.

Since the online effect was found in this task but not in the previous experiment 3 (Dual Task Self-paced Reading), the next experiment sought to identify whether the relatedness effect online was a task-based effect which is much easier to find in the maze task and very difficult (if at all possible) to find with self-paced reading. To test this, another experiment was run using a Simple Self-paced Reading Task.

2.8 Experiment 5: Simple Self-paced Reading Task

To test whether it was the demanding nature of the task that allowed for the relatedness effect to reach significance in online measures in the maze task, another simple self-paced reading task was run.

Based on the thematic integration account which suggests that increased demands lead to an observation of the effect, no significant effect of relatedness was expected, as self-paced reading is not a particularly taxing task which could allow for the effect to be found online. Therefore, similar reading times across conditions were expected for the main verb at the critical region. The effect however was expected to be found offline as in experiment 3 and in line with Van Dyke and McElree (2006) and Van Dyke et al. (2014) where self-paced reading was also used. The effect should be found both in response times to comprehension questions, with faster response times in the related condition overall, as well as in accuracy in comprehension questions, with the related condition predicted to have higher accuracy.

2.8.1 Participants

48 participants were recruited via the Prolific Academic platform. Subjects were native British English speakers with normal or corrected to normal vision, no literacy or hearing difficulties and no language disorders. Informed consent was provided and all participants were compensated for the completion of the experiment (£5.63).

2.8.2 Materials

The same materials (experimental sentences, fillers, comprehension questions) from experiment 4 were used in experiment 5 as well. The critical manipulation was Relatedness, with the two conditions being the related, where the main verb and the filler noun were more closely related (i.e., *boat, sailed*) and the unrelated where the filler noun and the main verb were not as closely related (i.e., *boat, fixed*). An example experimental item is given in table (2.8). Each sentence was followed by a comprehension question with an equal ratio of yes / no responses.

Relatedness	**Sentence**
related	It was the boat that the guy who lived by the sea **sailed** in two sunny days.
unrelated	It was the boat that the guy who lived by the sea **fixed** in two sunny days.

Table 2.8: Example of experimental item in simple self-paced reading task

All the experimental stimuli for the experiment can be found on p.169.

2.8.3 Procedure

The experiment was coded using the Ibex Farm software (Zehr and Schwarz, 2018b, Drummond, 2013) and was run via the Prolific Academic platform. Two counterbalanced lists were created so that each item appeared only once in one condition in each list. Participants read sentences word by word in a self-paced design. Each sentence began with a dash and participants could move to the next word by pressing the 'space' key on their keyboard. Each sentence was followed by a comprehension question. To respond to the question, participants were required to press the 'e' key on their keyboard if the correct answer was on the left side of the screen, or the 'i' key on the keyboard if the correct answer was on the right. The experiment lasted approximately 45 minutes.

2.8.4 Analysis

A threshold of 75% accuracy in responses to comprehension questions was required for inclusion in the analysis, leading to data from three subjects to be excluded. Reaction times below 100 ms or above 10.000 ms were also excluded from the analysis.

The levels of the factor Relatedness were contrast coded, using -1 (related condition), and 1 (unrelated condition). A high probability that the Relatedness parameter is smaller than zero would suggest that there is evidence that the related condition is processed faster and / or more accurately than the unrelated condition. By contrast, a high probability that the parameter Relatedness is greater than zero would mean that the condition where the verb is unrelated is processed more accurately / faster.

All other statistical analysis procedures and parameters were identical to those of experiment 3.

2.8.5 Regions of Interest

The critical and the spillover regions were the two regions that were analysed and will be reported here. The regions in this task are identical to the ones in experiment 4, with the critical region including the main verb of the sentence (i.e. *sailed / fixed*), and the spillover region including the first word at the position right after the main verb (i.e. *in*).

2.8.6 Results

Models are reported for offline and online measures: offline for response times and accuracy to comprehension questions, and online for reading times at the two regions

of interest: the critical region (verb: i.e., *fixed / sailed*) and the spillover region (word following the verb: i.e., *in*).

Reading Times

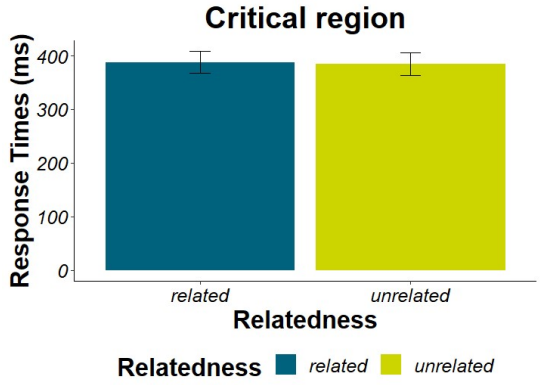


Figure 2.11: *Mean reading times at the critical region in experiment 5. Error bars represent the standard error of the means.*

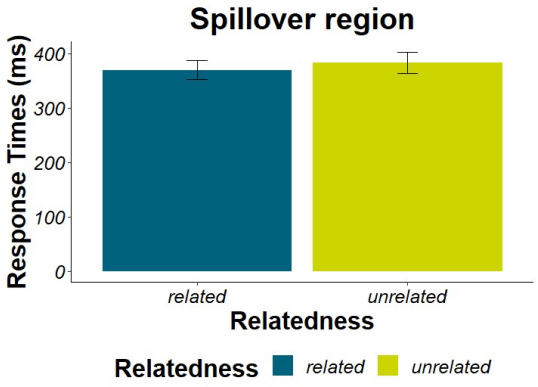


Figure 2.12: *Mean reading times at the spillover region in experiment 5. Error bars represent the standard error of the means.*

Condition	Critical region	Spillover region
Related	387.56 (20.60)	369.48 (17.46)
Unrelated	384.10 (21.01)	382.43 (19.59)

Table 2.9: *Mean reading times (ms) per condition by region in experiment 5. The SE is given in parenthesis.*

Average raw reading times by condition in the critical and spillover regions are reported in table (2.9). The mean response times by region are plotted in figures (2.11 & 2.12). The mean of the posterior distribution over the Relatedness effect in the critical and the spillover regions, along with the 95% HPDI, are summarised in table (5.9, p.139). The posterior distributions are plotted in figure (5.13, p.146) for the critical region and in figure (5.14, p.146) for the spillover region.

As expected, the model for the reading times in the critical region provided evidence for a lack of an effect of Relatedness. The posterior distribution for the effect of Relatedness showed a mean of 0 ($\hat{\beta}$= 0.00, 95% CrI = [-0.03, 0.03], P($\hat{\beta}$ > 0) = 0.52), suggesting that the experimental condition of Relatedness caused no change in the reading times of the verb.

The posterior distribution for the reading times in the spillover region showed that the credible interval does not include zero ($\hat{\beta}$= 0.03, 95% CrI = [0.00, 0.07], P($\hat{\beta}$ > 0) = 0.98), suggesting that there was an effect of Relatedness in the direction of the spillover word being read slower when following a verb in the unrelated condition.

Comprehension Question Accuracy

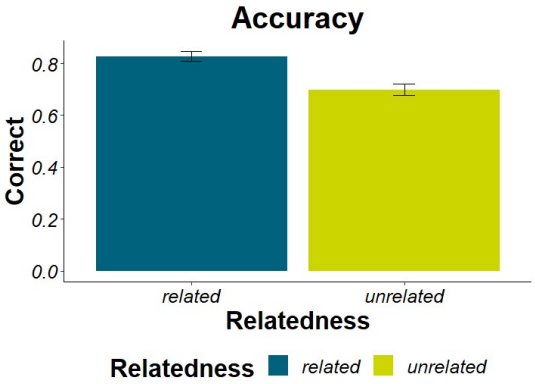


Figure 2.13: *Proportion of correct responses by condition in experiment 5. Error bars represent standard error.*

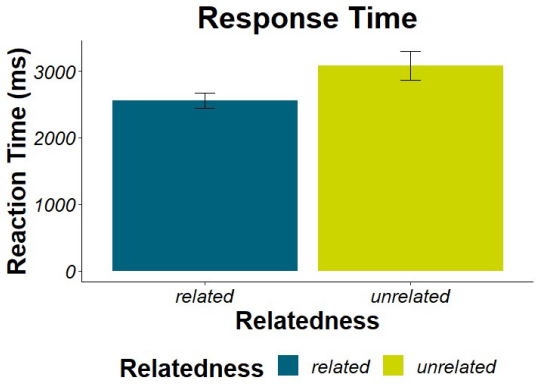


Figure 2.14: *Mean response times (ms) to comprehension questions in experiment 5. Error bars represent standard error.*

The mean percentages of correct responses in each condition are summarised in figure (2.13). Mean response times of correctly answered comprehension questions are shown in figure (2.14). Table (5.10, p.140) summarises the mean of the posterior distribution over the experimental fixed effect parameter and the 95% CI in the accuracy model and table (5.11, p.140) summarises the same for the response times model. The posterior distributions are plotted in figure (5.11, p.145) for the accuracy model and in figure (5.12, p.145) for the response times model.

The posterior distribution for the main effect of Relatedness provided evidence for a strong Relatedness effect. It has a range of negative values only ($\hat{\beta}$= 0.74, 95% CrI = [-1.43, -0.05], P($\hat{\beta}$ < 0) = 0.98), suggesting that accuracy in comprehension questions following sentences with a verb in the related condition was higher and subjects were not as accurate when answering questions following sentences with the verb in the unrelated condition.

Comprehension Question Response Times

The analysis of the comprehension question response times showed that the posterior distribution for the main effect of Relatedness provides evidence for a very strong Relatedness effect. Not crossing 0 and having a range of positive values only ($\hat{\beta}$= 0.12, 95% CrI = [0.06, 0.18], P($\hat{\beta}$ > 0) = 1), suggest that subjects were faster responding to comprehension questions following sentences with a verb in the related condition than questions following sentences with the verb in the unrelated condition.

2.8.7 Discussion

The analysed data of this task painted a very interesting picture with regards to the predictions set by the thematic integration account.

Consistent with the thematic integration account expectations and at the same time with the previous two experiments (experiments 3 and 4), in the offline data the relatedness effect was very strong. In the comprehension question responses, participants were consistently faster and more accurate when these questions followed sentences in the related condition compared to those appearing after sentences in the unrelated condition.

When it comes to the online data analysis, the initial predictions based on the thematic integration account were partially correct. Consistent with the account's expectations, there was no effect in the critical region of the sentences, which is the region consisting of the main verb. Overall, the response times for that region were very similar and relatedness did not seem to have an impact in that region. However, unlike what was initially expected in the online measures, a relatedness effect was found in the spillover region (i.e. *in*), with participants being faster in that region after having read the verb in the related condition over the one in the unrelated condition. Some potential reasons for this unexpected effect will be discussed in the General Discussion section of this chapter (2.1.9).

2.9 Effect of Task on Relatedness

To better explore the effect of Relatedness depending on the different tasks previously run (self-paced reading and maze tasks), a comparison model was run. In particular, the data of the two tasks that were in all but the type of task identical (experiments 4 and 5), were collapsed into a single model, comparing whether the different task could account for the difference in the effects found. The reaction times of the two tasks were combined in one data frame, with the same criteria used as in the tasks (i.e. exclusion of extreme values). They were then analysed with a Bayesian linear mixed effects model that assumes a Gaussian distribution, with the effects of Relatedness and Experiment, and their critical interaction. The levels of Experiment were coded as 1 (for experiment 4: the maze task) and -1 (for experiment 5: the simple self-paced reading task). For the fixed effects, normal priors were set, with a mean value of 0 and the variance of the distribution set to 1.

The analysis of the collapsed model apart from the expected Relatedness effect, also revealed an effect of Experiment, as well as a critical interaction between Relatedness and Experiment. The mean of the posterior distributions over the fixed effect parameters together with the 95% highest posterior density interval (HPDI) are given in table 5.12, p.140. The posterior distributions are plotted in figure 5.15 p.146. The posterior distribution for the effect of Experiment showed a very large mean ($\hat{\beta}$= -1.09, 95% CrI = [-1.13, -1.05], P($\hat{\beta}$ < 0) = 1), suggesting that in the Maze Task the reaction times overall are shorter than the ones in the Simple self-paced Reading Task. The posterior distribution for the expected effect of Relatedness showed a positive credible interval ($\hat{\beta}$= 0.09, 95% CrI = [0.04, 0.14], P($\hat{\beta}$ > 0) = 0.99), suggesting that the effect of Relatedness was much stronger in the Maze task than in the Simple Self-paced Reading task. The interaction between Relatedness and Experiment had a credible interval of negative values only ($\hat{\beta}$= -0.09, 95% CrI = [-0.14, -0.04], P($\hat{\beta}$ < 0) = 0.99).

The output of the model could account for the very strong effect of Relatedness in the critical region in the maze task (experiment 4) but the complete lack of an effect in the same region in the simple self-paced reading task (experiment 5). Overall, the model run here shows that the different experimental task might be the driving force behind the observed (or the lack of) Relatedness effects in particular regions of the tested sentences.

2.10 General Discussion

The findings from the last three experiments (Experiments 3 - 5) are consistent with the thematic integration account. The two self-paced reading tasks and the maze task show that the observed effects are due to the relationship between the filler noun and the verb and could not necessarily be attributed to similarity based interference.

2.10.1 Offline Effects

The offline results were very consistent across all three experiments, confirming the thematic integration predictions. Overall, accuracy in comprehension question responses was higher when responding to a question after a sentence where the verb was semantically related with the filler. These results are in line with the effect accuracy data in the original study (Van Dyke and McElree, 2006), where responses were more accurate in the non-interfering condition (which is the condition named as related in the current study , i.e., sentence with *fix*). Comprehension question response times followed the same pattern confirming the thematic integration account's expectations across all experiments. In all three experiments, there was evidence for very strong effects with subjects being faster in their responses to questions following sentences containing the verb in the related condition (i.e., *sailed*).

While the offline effects are in line with the expectations set by the proposed thematic integration account, they do not seem to go against the expectations set by the cue-based model either. Retrieval theories posit that retrieval is primarily a race between items, where each item competes to receive the most activation as quickly as possible (Cunnings & Sturt, 2018; Lewis & Vasishth, 2005; Nicenboim & Vasishth, 2018; Smith & Vasishth, 2020; Van Gompel et al., 2000). That appears to mean that cue based models make clearer predictions primarily for online processing, as that is when the sentence is being unfold in real time, when retrieval happens and when dependencies are restored. Offline measures normally test the interpretation assigned to a sentence, which appears to be a post retrieval measure rather than an indication of whether retrieval was difficult online in the case of retrieval interference.

2.10.2 Online Effects

The online effects in the last three experiments (experiments 3 - 5) were the primary focus of the present research. It is those online measures that allow for the direct testing of potential differences in reading times as an effect of retrieval interference or thematic integration. The online effects in the three experiments were partially predicted by the thematic integration account, with no effect in the dual task self-paced reading study (experiment 3), a strong effect in the critical region in the maze

task (experiment 4), and no effect in the critical region as well as an effect in the spillover region in the simple self-paced reading task (experiment 5).

In line with the thematic integration expectations, there was a strong relatedness effect in the critical region in the maze task (experiment 4). Subjects were faster to read and select the verb in the related condition than that in the unrelated condition. Forcing subjects to make correct selections of words seems to reinforce the online effect in the critical region as there is a faster selection when the filler noun and the verb are strongly related.

Unexpected (?) Effect

Also in line with the thematic integration account there was no effect in the critical region in the simple self-paced reading study (experiment 5), as the task was considered too easy for the effect to appear. However, a result not falling under the initial predictions, was the online relatedness effect in the spillover region in the same study. The original prediction for that study was the complete lack of online relatedness effects, expecting similar reading times in both conditions, both in the critical and the spillover regions. The relatedness effect was expected to be observed offline only. This is nevertheless not problematic for the understanding of the relatedness effect or for the thematic integration account.

Although the observed effect could easily be attributed to noise in the experimental data, it could instead be explained within the thematic integration account previously discussed. The thematic integration account suggests that when building dependencies, the parser focuses on the semantic relationship between filler and verb. This relatedness relationship is always present but it could be easier to observe or measure its effect when there is increased complexity.

The reason for the lack of an effect in the spillover region in the no memory load condition in Van Dyke and McElree (2006) and in Mertzen et al. (2020) is twofold. The spillover region in those studies consists of multiple words (i.e. *in two sunny days*). In Van Dyke and McElree (2006), the reported results refer to the four words following the verb as the spillover region, where only a memory load effect is found. Mertzen et al. (2020) report that no effects were observed in that same post-critical region for none of the languages they tested. The present simple self-paced reading task (experiment 5) was modelled after the maze task I (experiment 4), where words are presented one by one; so in the reported results as spillover region is considered just the one word following the verb (i.e. *in*). Importantly, in the present data this relatedness effect is only found at the one word in the post-verbal region and then it goes away pretty quickly. The effect might not have been detected then, due to the fact that multiple words were collapsed in one region only (Mertzen et al., 2020; Van Dyke & McElree, 2006).

Furthermore, the spillover region in the aforementioned studies is the final region at the end of the sentence. There are multiple studies in the literature suggesting that when testing the final part of a sentence, there are wrap up effects which could have potentially cancelled out the relatedness effect due to its size and its brief emergence (Balogh et al., 1998; Fodor et al., 1996; Hill & Murray, 2000; Just et al., 1982; Mitchell & Green, 1978; Rayner et al., 2000; Swinney & Zurif, 1995; Warren et al., 2009; White et al., 2011). As the relatedness effect seems to be rather sensitive and difficult to observe online, looking for it in a region where all those effects are at play due to the region's position in the sentence, minimises the probabilities of finding it.

So it can be hypothesised that even if the relatedness effect in those previous experiments could momentarily reach significance for one of the words, gets lost due to the post-verbal region's treatment as a sum of multiple words at the very end of the experimental sentence.

Based on that then, this unexpected result could also explain the lack of an online effect in the dual task self-paced reading study (experiment 3). The lack of an online effect in the critical and in the post critical regions of the sentence in that task could clarify how this effect really works online. The dual task self-paced reading experiment was a close replication of Van Dyke and McElree, 2006, with the main difference being the addition of a verbal memory load instead of the absence of one. That means that the regions in the study were identical. Despite the task being complex due to the constant memory load presence across all conditions, the self paced reading technique is not sensitive enough to measure a small effect like relatedness online, especially when it might be cancelled by other effects in a final region with multiple words.

2.10.3 Theoretical Accounts

Thematic Integration

From the three experiments it is clear that the relatedness effect is not only consistently present offline, but it is an effect that can be detected online too. However, it becomes evident that the magnitude and the timing of the effect is heavily dependent on the task and the manipulation of the experimental materials. In the maze task (experiment 4), which is a difficult and sensitive task which forces incremental parsing, the relatedness effect appears early in the critical region which is a region consisting of one word only. When testing with a different method, namely self-paced reading, which is a less sensitive paradigm than the maze the effect appears later in the sentence. It appears in the spillover region instead of the critical region, and that is only under certain conditions. In the simple self-paced reading task (experiment 5), despite the task not being particularly difficult, there is an effect in the spillover

region, possibly because it consists of one word only as in the maze task. However, in the dual task self-paced reading (experiment 3), the task which is more complex than experiment 5, there is no effect in the spillover region as it consists of multiple words, so the effect gets diluted.

The thematic integration account then needs to be redefined such that : in dependency resolution the parser does focus on the semantic relationship between filler and verb, and this effect can be seen both online and offline. However, despite this effect being robust and easy to observe offline, during online processing there need to be certain conditions. From the three experiments presented here, it is evident that the online effect is not only task dependent, being easier to see an effect when the task is more demanding, but also that the design of the experiment should be such that it would allow for the observation of such a fickle effect (i.e. smaller regions of interest).

Cue-based Model

Against the cue-based model predictions, the dual task self-paced reading study (experiment 3) showed that the contents of the memory load do not seem to interfere when forming a dependency between a filler and a verb, so there is no semantic attraction as it was originally suggested by Van Dyke and McElree (2006). There In the nominal memory load condition all three nouns could be plausible arguments of the clefted verb, whereas in the verbal load condition the three verb were always impossible arguments of the verb. The absence of an effect of memory load type could suggest that having a memory load at all regardless of its content might serve the same purpose with the maze task design (experiment 4); it provides increased difficulty to the task, which in turn forces the parser to focus on the lexical information of items to establish a dependency. That could lead to facilitation during processing thematic integration when the retrieval target filler noun and the verb are strongly related, which if testing with an adequate technique and the right design would be detectable.

Cue-based models predict that there should be interference effects when resolving dependencies and there are more than one fitting competitors. In the series of experiments run specifically on filler-gap dependencies here, that is not really seen. In particular in the dual task self-paced reading experiment (experiment 3) which actually involved potentially fitting competitor words, there was no sign of interference caused by those sentence external distractors; they appeared not to lead to semantic attraction effects. What was tested for though, was interference at the point of retrieval. Other studies have shown that in filler gap dependencies interference effects could be caused from sentence external distractors but at the stage of encoding, as

found in Gordon et al. (2002) and Fedorenko et al. (2006).

Final Conclusion

This first set of experiments showed that the observed results are best understood as stemming from purely sentence internal mechanisms, as there seems to be a prioritisation in retrieving items which are in structurally licensed positions when forming dependencies. There is no result in these experiments which would suggest that sentence external items interfere with sentence internal dependencies, as the cue-based model would predict, even when there is partial match between the cues of those items and the cues of the retrieval verb (i.e., the memory load items *table, sink, truck* being *+NOUN, +FIXABLE* among other things). The results suggest that when building a dependency the parser does indeed consider lexical information which can lead to a difference in processing times, but that is only between items with a structural relationship (direct object - main verb).

Overall, what the data suggest is that during the resolution of a filler-gap dependency the parser focuses primarily on the unique semantic relationship between sentence internal structural terms which is strongly manifested in offline results, and which depending on the manipulation of the sentences and the experimental technique might lead to facilitation during processing. Moreover, the dual task self-paced reading experiment (experiment 3) in particular, which tested for the effect of sentence external distractors clearly showed that there is no attraction effect during filler-gap dependency resolution that could be attributed to retrieval interference caused by sentence external distractors.

Chapter 3

3.1 Similarity based retrieval interference and semantic attraction from sentence internal distractor words

Quite different but equally as interesting as the previously presented set of studies in the last chapter (chapter 2), in this present chapter similarity based interference effects will be discussed in relation to grammatical illusions. This chapter focuses on a different type of observed effects and while chapter 2 explored semantic attraction as similarity based interference effects in filler-gap dependencies from sentence external distractor nouns, what will be investigated here is how interference effects from sentence internal distractor words might give rise to semantic illusions when it comes to the interpretation of dependencies in sentences with manipulated plausibility. Having as a starting point a study by Cunnings and Sturt (2018), this chapter is an attempt to aid understanding attraction phenomena slightly better.

3.1.1 Attraction effects

Multiple studies testing the predictions of cue-based architecture models (Lewis & Vasishth, 2005; McElree, 2000; McElree et al., 2003; Van Dyke, 2007), have focused on so called attraction effects (Bock & Eberhard, 1993; Bock & Miller, 1991; Eberhard, 1997; Pearlmutter et al., 1999; Tanner et al., 2014; Wagers et al., 2009). Attraction effects refer to instances in which the parser / generator appears to erroneously retrieve an incorrect item during either sentence processing or production. A classic example of attraction in agreement is shown in 24.

(24) a. The **key** to the **cabinet / cabinets** is/are.

Items in inaccurate position appear to be incorrectly retrieved because of their partial matching with one or more cues in that position and treat it as a complement of the verb that is the source of these cues. For instance in the given example in a production study, the percentage of the production of the incorrect *are* was higher

than that of the correct *is*, when the sentence contained the plural *cabinets*. That suggests that the inclusion of the plural form had an effect on the production of the verb, despite the pluralised noun not being in a position that should have informed the parser.

Another approach that has been adopted when it comes to effects like the ones found in sentences like 24, is the representation distortion accounts (Bock and Eberhard, 1993; Eberhard, 1997; Futrell et al., 2020; Yadav et al., 2023). Those are different from cue-based retrieval models, in that instead of items sharing features/cues and getting activated due to a match of those, those accounts adopt the idea that features (and especially the feature of number) float and percolate on other items. Accounts like those, assume that the feature of number of the subject noun (i.e., *key* in example 24) gets overwritten probabilistically by the number feature of the noun in the prepositional phrase (namely *cabinets*) which leads to an inaccurate memory trace of the *key*. Importantly, the feature of plurality is stronger than that of singularity. In cases of grammatical attraction then like the one in example 24, the plural feature of the distractor noun *cabinets* percolates and overwrites the singular feature of the subject noun *key*, leading to the subject noun *key* now being marked as plural and matching the verb *are* in number. That percolation of the plural feature then, makes the sentence containing *are* seem grammatical (illusion of grammaticality).

Importantly though, in cue-based retrieval frameworks those attraction effects have been suggested to be a strategy that the parser adopts to "rescue" sentences that are otherwise ill-formed, hence ungrammatical. Therefore it is not considered to be a default strategy that is chosen and followed across all types of sentences; but that attraction is what happens when sentences are not well built and would not allow for successful parsing (Laurinavichyute and von der Malsburg, 2022, Lago et al., 2015).

The asymmetry though between the existence of attraction effects in ungrammatical sentences and the claimed lack of such effects in grammatical sentences has been an issue of debate given findings from some studies that have given support to such an asymmetry (**i.e.**; Bock & Miller, 1991; Franck et al., 2002; Pearlmutter et al., 1999; Tanner et al., 2014; Wagers et al., 2009), while there are also findings that show that attraction can also be present in grammatical sentences (**i.e.**; Bock et al., 1999; Eberhard, 1997; Hartsuiker et al., 2003; Staub, 2009; Vigliocco et al., 1996). There are multiple studies that do not find any effects in grammatical sentences, and attraction effects from sentence internal distractor words are only observed when the tested sentences are ungrammatical. In such sentences (i.e., example 24) what is reported is that processing is facilitated in ungrammatical sentences (i.e., *The key to the cabinets are...*) due to the distractor noun (i.e., *cabinets*) matching some of the features of the verb (i.e., agreement in number feature, plural in the example 24) and therefore more easily retrieved, whereas no such match becomes available in grammatical sentences

as there is a match between the features of the actual subject noun and the verb, leaving little need for interference for match with any other term in the sentence (Dillon et al., 2013; Nicol et al., 1997; Pearlmutter et al., 1999; Tanner et al., 2014; Wagers et al., 2009). At the same time, both online as well as offline measures taken with a variety of methods have shown that in grammatical sentences too a distractor word could have an impact on processing. Studies have shown that an intervening distractor noun could impact processing in that it could lead to delayed processing when there is partial match of features between that distractor and a verb in the sentence (Häussler, 2009; Patson & Husband, 2016; Pearlmutter et al., 1999; Tanner et al., 2014).

One particular case with a set of findings offering support to this grammatical asymmetry is Cunnings and Sturt (2018).

3.1.2 Illusions of grammaticality

Cunnings and Sturt (2018) investigated semantic attraction effects in sentences where the plausibility of the whole sentence and the plausibility of intervening distractor words were manipulated, with each experimental sentence containing two different nouns that could potentially have cues that would match those of a verb in the tested sentences. The plausibility of the sentence as a whole as well as that of an intervening distractor noun were manipulated, such that there was a match or a mismatch between the grammatical noun and / or the distractor noun and the verb in the relative clause modifying the grammatical noun in sentences like that in table 3.1.

Sentence	Distractor	Sentence
plausible	*plausible*	Sue remembered the **plate** that the butler with the **cup** accidently shattered today in the dining room.
	implausible	Sue remembered the **plate** that the butler with the **tie** accidently shattered today in the dining room.
implausible	*plausible*	Sue remembered the **letter** that the butler with the **cup** accidently shattered today in the dining room.
	implausible	Sue remembered the **letter** that the butler with the **tie** accidently shattered today in the dining room.

Table 3.1: Example experimental items in Cunnings and Sturt, 2018

Sentences like the above were tested in an eye-tracking while reading paradigm, recording several measures, both early and late. The cue-based retrieval models (Lewis & Vasishth, 2005; McElree et al., 2003; Van Dyke, 2007) predict that when there is a match between the cues of an element in the sentence and those at the verb site, processing should be affected. Considering that attraction effects are more (if not only) prominent in implausible sentences compared to their plausible counterparts based on claims of studies previously run, attraction should lead to differences

between the implausible sentences when the distractor was plausible compared to when the distractor was implausible. That should have been manifested in facilitated fixation times at the critical region of the verb in the implausible sentences (sentences containing the noun *letter*) with a plausible distractor (so when *shattered* was preceded by *cup*) compared to more increased fixation times in the same region for the implausible distractor (when *shattered* was preceded by *tie*).

The results of that study as expected in both early measures (first pass and regression path times) and late measures (total viewing times) showed a main effect of the plausibility of the sentence, with participants having longer overall reading times for the implausible than the plausible sentences. Importantly, a significant interaction between the plausibility of the distractor and the plausibility of the sentence in total times ($p<.05$) was observed, with the implausible sentences having shorter reading times when the distractor was plausible compared to when the distractor was implausible, but no such effect in the plausible sentences. That is sentences containing *letter...shattered* had overall shorter total reading times for *shattered* and for *today* when those sentences contained the word *cup*, compared to when they contained the word *tie*. Those results were taken to as support to the cue-based retrieval model predictions, as they are considered to indicate an illusion of plausibility. The effect is taken as compatible with the model suggesting that the processing of the implausible sentences was facilitated in the presence of distractor words that were potentially plausible arguments of a verb in those sentences, by partially sharing some of the features that the complement of the verb should have.

3.1.3 Present set of studies

What is of great interest given those findings was attempting to understand whether they truly support the cue-based retrieval models or if the observed effects could be attributed to or explained better in a different way (i.e., due to the structure of the sentence, or due to complexity). Perhaps a potential initial question that arose given the presented results in Cunnings and Sturt, 2018 was whether the reported facilitation in total fixation times is due to the ambiguous nature of the verb in the relative clauses. The verb in the relative clause (i.e., *shattered*) could either be easier to retrieve as the verb of the clause when preceded by something that could match some of its features, or the facilitation during its processing could be attributed to it being locally coherent with the distractor noun. That would suggest that the observed facilitation in processing might be due to the parser interpreting the combination of the verb (i.e., *shattered*) and the preceding distractor word when it is a plausible complement for it (i.e., *cup*) locally, as a reduced relative clause (i.e., *... the cup* ***that was*** *accidently shattered...*). Something which could of course not lead to an effect that

can be considered as an illusion of plausibility, but purely an effect arising locally in the sentence.

The present set of studies was build based on three different hypotheses. All of them attempted to look into the semantic attraction effects reported in the previous study and truly test whether other factors could account for the observation of such an attraction of the verb by intervening distractor words: 1. ambiguous syntactic structure of the sentence, 2. implausibility of the sentence, and 3. sentence complexity.

The very first question to address with the following set of experiments was to investigate if the attraction effects reported in Cunnings and Sturt, 2018 could be explained via the particular structure used in the experimental sentences which could be considered as ambiguous. In essence, that is whether the shorter fixation times are due to the attraction of the verb by the distractor noun or because participants interpreted it as a modifier of the preceding distractor noun (i.e., *cup* potentially plausible match for *shattered*, vs. *...cup (that was) shattered...*).

Then the focus would go to implausibility as a potentially essential trigger for the observation of such attraction effects, as in this particular study attraction effects were found in implausible sentences only. Hence, the second topic of investigation was whether attraction effects are indeed a strategy the parser adopts in order to rescue, even partially, the meaning of a sentence.

And the last issue to explore was whether it was complexity that gives rise to the observation of attraction effects. Importantly, if structural complexity overloads the parser and perhaps leads to attraction effects due to the difficulty in extracting meaning out of difficult to form dependencies.

3.2 Experiment 6: Plausibility Study for Maze Task II

Experiment 7 sought to investigate whether there was any basis on this initial concern that the reported results in Cunnings and Sturt (2018) are due to the ambiguity that is associated with the way the critical verbs are presented or if the findings could be attributed to attraction by a semantically plausible distractor word.

To test this first question set for this batch of experiments, the ambiguity in the structure of the sentences had to be resolved. The maze task became the task of choice with the critical verb used either as an unambiguous past participle or as an ambiguous past tense that could serve either as the main verb of the relative clause in the tested sentences, or as a modifier for a matching intervening distractor noun. For consistency, it was essential to ensure that all the verbs tested in this experiment had the same form both in past participle as well as in the past tense. As some of the verbs used in the previous study did not fulfil those characteristics having different forms as a past tense which is what was used in the study and as a past participle (i.e., *drove, wrote, ate* vs. *driven, written, eaten*), they had to be eliminated and replaced with different verbs. A plausibility study (experiment 6) was essential for the additional experimental materials and it was set to test those extra verbs as well as new experimental sentences and their fit with the already tested sentences in Cunnings and Sturt (2018).

3.2.1 Participants

24 native British English speakers were recruited via the Prolific Academic platform (www.prolific.com) for this study (14 females - 10 males, median age: 32.1). All participants self-reportedly had no literacy or language related disorders. Informed consent was required before participation and compensation for participation was given after the completion of the study (£1.50).

3.2.2 Materials

The experimental items from Cunnings and Sturt (2018) were modified in such way so that conditions involving both a past tense or a past perfect would have the exact same from of the verb, as in the original study there were some sentences that included verbs with anomalous tenses (i.e., *Pauline saw the box that the man by the car/bin very carefully drove despite all the bad traffic.*).

Unlike the Cunnings and Sturt (2018) study, where the plausibility of the sentence is manipulated and tested, a decision was made to test implausible sentences only. In the original study, there was no effect found in the plausible sentences, where both the retrieval target and distractor noun could be plausible arguments of the verb. In

experiment 7 the focus fell on the illusion of plausibility effects that are found in Cunnings and Sturt (2018) implausible sentences only, when there is a manipulation of the plausibility of the distractor noun.

One of the manipulations in the present study was the ambiguity of the verb. The ambiguous condition is taken to be a verb in past tense as in the original study; a state which allows for a dual interpretation of the verb as either a past tense verb or as a past participle. The unambiguous condition involved the verb in past perfect. For consistency, the main criterion in the verb selection was that the verb in both the ambiguous and the unambiguous condition should have the same form. Of the original verbs in the experimental sentences in Cunnings and Sturt (2018) only 21 matched this criterion. Therefore in the remaining 11 the verbs were changed, as the form of the past participle and past tense were different.

9 new sentences following the same syntactic pattern were also created anew, in which the forms of the past participle and the past tense of the verb also had to match. In those new sentences, caution was taken so that the chosen verbs would be closely semantically related with the plausible nouns used in those sentences (both the retrieval target noun and the distractor noun) to allow for testing retrieval effects based on semantic / lexical cues, as with the rest of the original sentences.

The sentences were tested in a design that involved 4 variants of each sentence, where each condition involved only one of the grammatical object nouns or the distractor nouns as the main grammatical object of the verb of interest. The verb in the sentences in this study was only tested in its past form, as it was only plausibility that was tested here, so having both versions of past tense and past perfect would not offer any further insight but simply complicate the design. As such, an example sentence in the study would like the following (3.2):

C	**Sentence**
1	Pauline saw the van that the man very carefully parked despite all the bad traffic.
2	Pauline saw the car that the man very carefully parked despite all the bad traffic.
3	Pauline saw the box that the man very carefully parked despite all the bad traffic.
4	Pauline saw the bin that the man very carefully parked despite all the bad traffic.

Table 3.2: Example of experimental item in experiment 6

18 items involving the combination of the grammatical subject followed by the prepositional phrase which included the critical distractor in the relative clauses in the newly formed sentences were also tested, having 2 conditions each, one for each type of distractor (3.3):

As fillers, an equal number of sentences with a structure similar to the one in the experimental items was used (total of 28 sentences). Of those, half (14 sentences) were highly plausible, 7 were highly implausible and 7 were used as a control that would

C	Phrase
1	the child with the flute
2	the child with the hat

Table 3.3: Example of experimental item in experiment 6

place them somewhere on the middle of the plausibility scale. To counterbalance the 18 plausible phrases tested, an equal number (18) of implausible phrases with the same structure (noun followed by a prepositional phrase) were also included.

3.2.3 Procedure

The study was first coded using the PC Ibex platform (Drummond, 2013; Zehr & Schwarz, 2018a; Zehr & Schwarz, 2018b) and it was then run online via the Prolific Academic platform (www.prolific.co). Participants were asked to make judgements of the plausibility of the sentences and the phrases that were presented on screen, by rating them in a scale of 1 - 7 with 7 being the most plausible and 1 the least plausible. 3 sentences with structures different to the one used for the experimental items were used for practice before the start of the task. The session was divided in two parts, with the first including the presentation of the experimental sentences and the relevant sentence fillers, whereas the second part was for the phrases, both experimental and fillers. Participants were informed of those two separate parts in the task and were notified before the start of the second part. Before starting the task, participants were encouraged to use the whole scale of 1 - 7 rather than judge all sentences and phrases as equally plausible or implausible.

3.2.4 Analysis & Results

Plausibility data were analysed in R (Team R Core, 2020), such that for each of the four condition of the experimental sentences and the two conditions of the experimental phrases, individual ratings per item were calculated as well as an average mean for each condition. The mean by condition is given on table 3.4 :

Condition	Mean
S-1	5.82
S-2	5.65
S-3	2.10
S-4	2.19
P-1	6.85
P-2	6.74

Table 3.4: By condition mean ratings in experiment 6

The mean ratings of the fillers were used to check that participants paid attention and gave accurate judgements. The analysis of the mean ratings for the fillers confirmed that the quality of the experimental data was adequate.

As the tested experimental sentences and phrases were more than necessary for inclusion in the next task, the choice between them was based on the by-item average rating, with items chosen based on how much higher they were when needed to be included in the plausible condition and how much closer to 1 they were to be used as the implausible condition.

3.3 Experiment 7: Maze Task II

The run of the plausibility study (experiment 6) allowed for the main experimental task to be set in order to investigate the initial question of whether the results in Cunnings and Sturt (2018) are due to the ambiguous nature of the tested critical verb or if they could be due to attraction from the closely matching distractor noun.

3.3.1 Participants

48 native British English participants (median age: 32.1, 32 females - 16 males) were recruited via the Prolific Academic platform to complete this task (www.prolific.co). Based on their pre-screening answers, participants had corrected or corrected to normal vision, no hearing impairments and no literacy or language disorders. Before participation in the experiment, informed consent was provided by all subjects and a compensation was given after the completion of the task. As the maze task is sometimes deemed as unnatural and due to the long length and complexity of the experimental sentences a decision was made to compensate participants with £7 as an incentive for their effort.

3.3.2 Materials

The original stimuli from the Cunnings and Sturt (2018) study together with the newly tested sentences from experiment 6 (plausibility study) were modified in a a 2x2 design, manipulating the plausibility of the distractor noun, and the ambiguity of the verb (table 3.5). The plausibility of the distractor noun manipulation created two conditions, plausible and implausible, whereas the ambiguity of the verb manipulation created the ambiguous and the unambiguous conditions.

The original stimuli from Cunnings and Sturt (2018) were modified and used in this task. Some of the sentences in Cunnings and Sturt, 2018 were just slightly modified such that they include the new manipulations only, whereas verbs in some of the stimuli sentences were replaced by the newly normed verbs which matched in their past and their participle form from experiment 6 (plausibility study).

The main differences between this task compared to the task in Cunnings and Sturt (2018), apart from the obvious difference of the task itself (this study — maze task vs. Cunnings and Sturt (2018) — eye-tracking) were two of the manipulations. First, as previously discussed the critical effect was observed in implausible sentences only and therefore only this type of sentences were used in the present task. The manipulation of the plausibility of the sentence that was critical in Cunnings and Sturt (2018) was not used here, with all the critical sentences being implausible. As there was no predicted effect in the plausible sentences, they were altogether removed,

which importantly allowed for more observations per participant for each of the conditions. The second difference was the manipulation of the ambiguity of the verb. In the eye-tracking study of Cunnings and Sturt (2018) the verb was always in simple past tense which could be potentially ambiguous. Here the verb was manipulated, such that it was either ambiguous or unambiguous, by manipulating whether the simple past form of each verb was used in the relative clause (i.e., *shattered*), or if it was disambiguated by the inclusion of the word *had* which allowed for one of the two alternative readings of the sentences only (i.e., *had...shattered*), as seen in table 3.5.

The manipulation that is constant in both this and the Cunnings and Sturt (2018) experiments was the critical manipulation of the plausibility of the distractor noun, which was either a plausible argument for the following verb (i.e., *cup...shattered*) or an implausible argument (i.e., *tie...shattered*) seen in table 3.5.

Distractor	Verb	Sentence
plausible	*unambiguous*	Sue remembered the letter that the butler with the **cup had** accidently **shattered** today in the dining room.
	ambiguous	Sue remembered the letter that the butler with the **cup** accidently **shattered** today in the dining room.
implausible	*unambiguous*	Sue remembered the letter that the butler with the **tie had** accidently **shattered** today in the dining room.
	ambiguous	Sue remembered the letter that the butler with the **tie** accidently **shattered** today in the dining room.

Table 3.5: Example of experimental item in experiment 7

As the maze is an already more demanding, relatively unnatural and taxing task, two critical changes had to be made. First, the experimental items were simplified by removing all the unnecessary information that might have been useful in the original Cunnings and Sturt (2018) eye-tracking design. All the introduction and wrap-up sentences were removed, as they were not really needed in this design, as well as the comprehension questions following each sentence, as comprehension is essential while a subject is tested on the experimental sentence considering that the task itself forces incremental parsing. The degree modifiers (i.e., quite / very / rather) preceding the adverb separating the distractor noun and the verb in some of the sentences in Cunnings and Sturt (2018) were also removed, but the adverb was retained so that the distractor noun and the verb were not adjacent. For consistency, the endings of the sentences were also trimmed or slightly modified, so that all sentences had around three words after the word that was the spillover region of the sentences.

And secondly, due to the difficulty of the maze task in comparison to the eye-tracking paradigm that was originally used in Cunnings and Sturt (2018), the ratio of experimental items to fillers was also changed from 1:3 to 1:2 (original study (Cunnings and Sturt, 2018): 32 experimental items, 96 fillers vs. maze task: 40 experimen-

tal items, 80 fillers). As the new design involved 40 implausible experimental items, 40 of the fillers were decided to be structurally similar but all plausible sentences, to counterbalance the plausibility effect and therefore to avoid for the experimental items to be easily identifiable. The plausibility of those 40 sentences was ensured as most of those fillers had also been tested in the plausibility study before this task (experiment 6). The rest of the fillers were all plausible sentences but with a different syntactic structure and varying levels of difficulty.

Foils for the experimental items and fillers were automatically generated with the A-maze software. Foils were manually checked and occasionally altered to avoid having two plausible continuations for the sentence, in particular for the regions of interest (i.e., *shattered* and *in* in the example experimental item in table 3.5).

All the experimental stimuli for the experiment can be found on p.175.

3.3.3 Procedure

The experiment was coded with the PC Ibex Farm software (Drummond, 2013; Zehr & Schwarz, 2018a; Zehr & Schwarz, 2018b), while the recruitment of participants was run via the Prolific Academic platform (www.prolific.co). To ensure that the stimuli appeared only once during the course of the experiment, four counterbalanced lists were created containing only one condition of each stimulus sentence. The procedure followed in this task was very similar to the one in experiment 4, with the main difference being that there were no comprehension questions following the experimental sentences, but simply a presentation of words where participants were required to make choices for each position in the sentence. The experiment lasted approximately for 40 minutes.

3.3.4 Analysis

To avoid the impact of extreme response times, all response times above 10.000 ms were excluded from the data prior to the analysis. Trials in which participants had selected the incorrect foil in any of the positions prior to the position of the critical region were also altogether excluded from the analysis.

Three Bayesian linear mixed effects regression models were fitted to the data, with the factor Distractor, the factor Verb and their interaction (Distractor$\times$Verb) treated as fixed effects and the number of Subjects and Items treated as random effects. The levels of the factor distractor were coded as -1 (plausible) and 1 (implausible), while the factor verb was coded as 1 (unambiguous) and -1 (ambiguous).

All other statistical parameters for the models and the analysis were identical to the ones in experiment 3.

3.3.5 Regions of Interest

For the present experiment, two regions will be reported, as relevant effects (if there were any) were expected to be observed primarily in these two : the critical and the spillover region. Following the design of the sentences in the original task in Cunnings and Sturt (2018) very closely, the critical region where some effects were observed remained the same, meaning that the critical region in the task was the verb in the relative clause (i.e., *shattered*). The task being a maze task suggests that words are presented one by one to the participants. The spillover region then was considered to be the one word that immediately follows the verb of the relative clause (i.e., *today*).

3.3.6 Results

Three models will be reported here: two models for the response times in the critical (i.e., *shattered*) and the spillover regions (i.e., *today*) and a model on the accuracy in making the correct choice of word at the critical region (i.e., *shattered*).

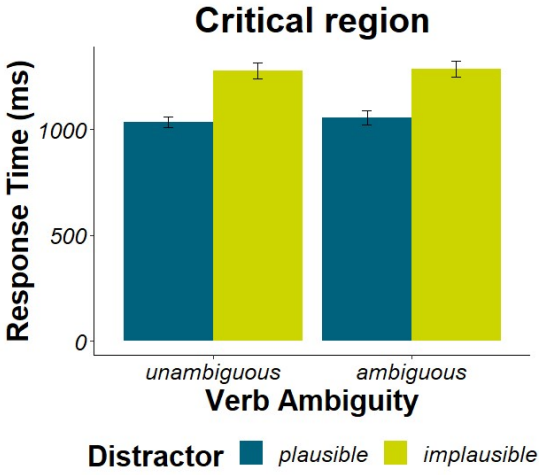


Figure 3.1: *Mean reading times at the critical region in experiment 7. Error bars represent the standard error of the means.*

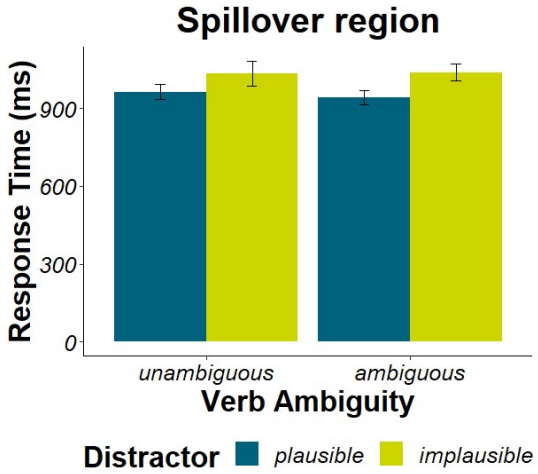


Figure 3.2: *Mean reading times at the spillover region in experiment 7. Error bars represent the standard error of the means.*

Condition	Critical region	Spillover region
Ambiguous/Implausible	1285.63 (39.35)	1039.11 (32.50)
Unambiguous/Implausible	1278.48 (36.93)	1035.56 (47.74)
Ambiguous/Plausible	1053.60 (33.4)	941.33 (28.18)
Unambiguous/Plausible	1036.25 (24.81)	963.40 (29.43)

Table 3.6: *Mean response times (ms) per condition by region in experiment 7. The SE is given in parenthesis.*

Response Times

The average raw response times by condition in the critical and spillover regions are given in table (3.6) and plotted in figures (3.1 & 3.2) for each region of interest. The standard error is given in the parentheses. The mean of the posterior distributions over the experimental fixed effect parameters in the critical and the spillover regions, along with the 95% highest posterior density interval (HPDI) are summarised in table (5.14, p.140). These Bayesian credible intervals indicate where the most plausible parameter values for these fixed effects parameters lie, given the data. The posterior distributions are plotted in figure (5.16, p.147) and in figure (5.17, p.147) for the critical and the spillover regions respectively.

The analysis of the response times at the critical region showed that the posterior distributions provided evidence for a main effect of the distractor plausibility manipulation, but they were otherwise uninformative about the other fixed effects. Specifically, the credible interval of the effect of the Distractor Plausibility manipulation spanned over positive values only, with an estimate of ($\hat{\beta}$= 0.09, 95% CrI = [0.07, 0.12], P($\hat{\beta}$ > 0) = 1), suggesting faster response times for the sentences containing a plausible distractor compared to slower response times for the sentences with an implausible distractor. However, the credible intervals of the effect of Verb Ambiguity, as well as the interaction of Verb Ambiguity and Distractor Plausibility spanned a range of values below and above zero, with an estimate for the Verb Ambiguity of ($\hat{\beta}$= -0.01, 95% CrI = [-0.03, 0.01], P($\hat{\beta}$ > 0) = 0.79) and for the Verb Ambiguity x Distractor Plausibility interaction ($\hat{\beta}$= -0.00, 95% CrI = [-0.02, 0.02], P($\hat{\beta}$ < 0) = 0.55).

The posterior distributions for the main effects in the spillover region were overall similar, albeit attenuated relative to the ones in the critical region. There was still some evidence for a Distractor Plausibility effect as the credible interval spanned over positive values only ($\hat{\beta}$= 0.04, 95% CrI = [0.01, 0.06], P($\hat{\beta}$ > 0) = 0.99), suggesting faster response times in the region in sentences containing a plausible distractor rather than an implausible one. The posterior distributions for the Verb Ambiguity and the interaction between Verb Ambiguity and Distractor Plausibility showed no such effect: Verb Ambiguity ($\hat{\beta}$= 0.00, 95% CrI = [-0.02, 0.02], P($\hat{\beta}$ > 0) = 0.56), Verb Ambiguity x Distractor Plausibility ($\hat{\beta}$= -0.00, 95% CrI = [-0.02, 0.02], P($\hat{\beta}$ < 0) = 0.61).

Accuracy

As accurate were considered the word choices that participants made for the critical region position that were the correct continuation of each tested sentence. Figure (3.3) summarises the mean percentages of correct responses in each condition. The mean of the posterior distribution over the fixed effect and the 95% CI in the accuracy model are summarised in table (5.13, p.140). The posterior distribution of the effects for the accuracy model are plotted in figure (5.18, p.147).

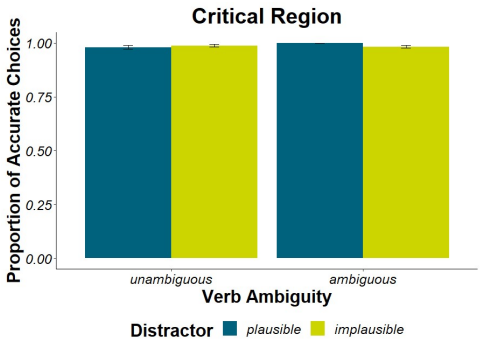


Figure 3.3: *Proportion of correct responses by condition in experiment 7. Error bars represent standard error.*

The posterior distribution for the fixed effects in the accuracy model provided evidence for a strong effect of the interaction between Verb Ambiguity and Distractor Plausibility. It has a range of positive values only, with an estimate of ($\hat{\beta}$= 0.87, 95% CrI = [0.03, 1.83], P($\hat{\beta}$ < 0) = 0.98), suggesting that a plausible distractor in sentences containing an unambiguous verb led to lower accuracy whereas a plausible distractor in sentences with an ambiguous verb led to higher accuracy in selecting the critical verb. The estimates of the effects of Verb Ambiguity and Distractor Plausibility showed no reliable effect, being ($\hat{\beta}$= -0.68, 95% CrI = [-1.67, 0.18], P($\hat{\beta}$ < 0) = 0.93) and ($\hat{\beta}$= -0.58, 95% CrI = [-1.50, 0.27], P($\hat{\beta}$ < 0) = 0.91) respectively.

3.3.7 Discussion

The results of this task were similar to the ones in Cunnings and Sturt (2018). The ambiguity manipulation that was tested here was not given any support by the data collected. The data here suggested that there could be attraction effects associated with the presentation of a plausible distractor word in sentences like the tested ones and a potentially plausible verb regardless of the ambiguity in the presentation of such a verb. The tested manipulation of a potential ambiguity that stems from how a verb was presented in Cunnings and Sturt (2018) (so when that said verb allows for two potential interpretations) appears to have no effect in whether semantic attraction effects will be detected. From the present results, it seems that the overall ambiguity of the verb phrase does not really affect whether there will be any semantic attraction effects at the verb at all, while the plausibility of the distractor appeared to be the only manipulation that had an effect on the speed with which participants made a choice between the two presented word options in the task.

Both in the critical region (verb) and the spillover region (word following verb, i.e., *today*), response times were modulated by the plausibility of the distractor noun and whether that said noun could act as a plausible complement for the verb. The

manipulation of the ambiguity of the verb did not show up in any of the speed data collected. That could suggest that the possibility that the found results in Cunnings and Sturt (2018) are due to the nature of the verb used in the tested sentences is not given support from this study. The disambiguation of the verb phrase with *had* as well as the distance that that word creates between the distractor noun and the verb appeared to play no role in the facilitation of processing the verb.

However the accuracy data, namely how accurately participants were in choosing the correct verb for the tested sentences, provided a slightly more interesting picture. Overall accuracy was very high across the board for all manipulations, but that data showed that in sentences with a plausible distractor noun, it was verb ambiguity that could lead participants to make a more accurate choice. For sentences with a plausible distractor noun that had an unambiguous verb, (so i.e., *...cup had accidentally shattered...*) participants were slightly less accurate in their choice of word at the critical region (so choosing the correct verb (i.e., *shattered*)) than sentences with a plausible distractor noun but an ambiguous verb (i.e, *... cup accidentally shattered...*). That could suggest that the ambiguity manipulation implemented here could on some level affect the way that participants make a choice. The added distance that the inclusion of unambiguous verb form allows for, might on some level alleviate the attraction effect when it comes to accuracy. However, from the present accuracy data no definite conclusion can be drawn as to whether it is the ambiguity of the verb in itself or the longer distance or even the choice of task that creates this effect.

Most importantly, the results of this present study showed that the Cunnings and Sturt (2018) results can be replicated with the use of another method, the maze task. Attraction effects can be found by employing a task not previously as commonly used to detect these effects. That then, allowed for a further investigation into these phenomena using this same task, as it showed reliably and quite confidently similar results with the ones from Cunnings and Sturt (2018) where the methodology is rather different (eye-tracking) and much more widely accepted and trusted in the effects that get detected. More specifically, the present results from the maze task appear to be much more comparable to those from later rather than early eye-tracking measures. They seem to be in agreement with the data that Cunnings and Sturt (2018) got in the total time reading times, as that is when the illusions of plausibility were observed in eye-tracking.

The question that arose from these data was whether the found attraction effects were purely because all of the tested sentences were implausible and that was what the next study explored.

3.4 Experiment 8: Maze Task III

Following the results of experiment 7 and having established that the maze task is a method which could give reliable results for attraction phenomena, similar to ones collected with other methods, this task was built to test whether attraction simply arises in implausible sentences as a function of the parser to repair otherwise problematic sentences and make some sense, even partial, of an otherwise implausible sentence. If then semantic attraction phenomena are the parser's attempt to partially repair implausible sentences, then these should only arise in this particular type of sentence, while if presented with a globally plausible sentence there should be no such repair as there would be no need for any.

So this task was built such that participants would come across plausible sentences but with a possibility of implausibility based purely on their own choices. In this maze task they would always be presented with one possible word that allowed for building a completely plausible sentence, but there would also be the possibility that a verb presented as a foil could potentially match with a distractor word in that sentence. What was of interest was whether participants would choose to build the sentence in such a way that it would end up being implausible.

Cue-based retrieval accounts would predict that in such a task participants should choose the words that would lead to an overall plausible sentence. If attraction effects are a manifestation of an attempt to repair something broken, when being in a position to build a sentence and in particular when being presented with words that allow for building a plausible sentence, attraction effects should not be present.

3.4.1 Participants

96 participants were recruited in the Prolific Academic platform to complete this task (www.prolific.co) (49 females - 47 males, with a median age of 33.3). All participants were native British English speakers with no visual or auditory impairment and no literacy or language disorders, based on their self-reported answers in their pre-screening questionnaire. Before their participation in the task, all participants were asked to provide informed consent and they were compensated for their time after the end of the experiment (£5.63).

3.4.2 Materials

Materials from experiment 7 were modified in a design manipulating the plausibility of the distractor noun. The 40 experimental sentences involved the same probes with the ones used for experiment 7 but with one difference that was deemed essential for the task. The experimental items for this task finished right after the presentation of

the verb in the relative clause. The relative clause verbs that were the correct choice for the sentences (i.e., *written* in the example experimental sentence in table 3.7), were verbs that would always allow for the noun in the main sentence to be a plausible complement, which was indeed the verb's true grammatical complement (i.e., *letter* in the example sentence given in table 3.7), while neither of the two nouns in the relative clause would be a good match as a complement for that verb (i.e., the two nouns *cup / tie* in the example experimental item in table 3.7)

Distractor	Sentence
plausible	Sue remembered the letter that the butler with the **cup** had recently **written / shattered** (foil).
implausible	Sue remembered the letter that the butler with the **tie** had recently **written / shattered** (foil).

Table 3.7: Example of experimental item in experiment 8

Since in the previous experiment only implausible sentences were tested and the verbs for this task needed to allow for plausible continuations, some of the verbs used here were taken from the original study by Cunnings and Sturt (2018), while others were verbs already in experiment 6, or some were altogether newly generated for the purposes of this study. The verbs that were used as foils here were the same verbs that were used in experiment 7, as they allowed for a globally implausible sentence. Those verbs used as foils were a plausible match for the noun in the relative clause in one of the conditions of the experimental items (i.e., the verb *shattered* being a plausible match for the *cup* in the relative clause in the example sentence in table 3.7), while it was not a good match neither for the noun in the relative clause in the other condition (i.e., *shattered* $\neq$ *tie* in the experimental item in table 3.7), nor for the noun in the main clause (i.e., *shattered* $\neq$ *letter* in the experimental item in table 3.7).

For consistency and since some of the continuations were not a good match for the adverb that preceded them, some of the adverbs were changed so that they were an equally good match for either verb in the experimental sentences. For instance, where in the same example sentence in experiment 7 there was *accidentally* (i.e., in the example experimental sentence in table 3.5), in this experiment *accidentally* would not be a good match for the verb *written* (i.e., in the example experimental sentence in table 3.7). Accordingly, the word *accidentally* was changed to the word *recently* which is an equally good match to both options *written* and *shattered*, while it also allows for equal meaning for both. Participants were made aware of the end of the sentence as a full stop was included in the presentation of the final word which incidentally was the critical word of this experiment. A full stop was included for both options participants were presented with, both the critical word (i.e., *written*) and the foil word (i.e., *shattered*).

All the experimental stimuli for the experiment can be found on p.187.

3.4.3 Procedure

The PC Ibex Farm software (Drummond, 2013; Zehr & Schwarz, 2018a; Zehr & Schwarz, 2018b) was used to code this task, while the coded experiment was then run via the Prolific Academic platform (www.prolific.co). Two counterbalanced lists were created, one for each of the conditions of the experimental stimuli, to ensure that only one condition of each item was presented to each participant. The design of the task followed the same steps with the ones in experiment 4, while once again as in experiment 7 no comprehension questions to follow the experimental sentences were used. The experiment lasted for approximately 30 minutes.

3.4.4 Analysis

Response times above 10.000 ms were removed from the data before analysing them, as they were considered extreme for this task. In order for a trial to be included in the analysis, participants were required to have only correct selections of words for all positions up to the critical region position. As the critical region position was indeed the final position for this experiment, that meant that participants had to make accurate choices for all positions in the sentences they were presented with.

Two Bayesian linear mixed effects models were fitted to the data, the accuracy model with the factor Distractor Plausibility and the response times model with the factors Distractor Plausibility and Verb Choice as well as their interaction (Distractor Plausibility $\times$ Verb Choice) treated as fixed effects in the models and Subjects and Items as random effects. Both Distractor Plausibility and Verb Choice had two levels, with Verb Choice coded as 1 (accurate) and -1 (inaccurate), whereas Distractor Plausibility was coded as -1 (plausible) and 1 (implausible).

The statistical models had the same statistical parameters with the ones in experiment 3.

3.4.5 Regions of Interest

Due to the nature of this task, there was only one region of interest - the critical region - which will be the one reported and analysed here. The critical region involved the verb in the relative clause (i.e., *written* in the example sentence in table 3.7), which was the final word of the stimuli sentences. In contrast to the preceding task (experiment 7), there was no spillover region in this task as the words in it were altogether removed in the stimuli to allow for the particular design with the inclusion of the experimental foil word.

3.4.6 Results

Two models will be reported for this experiment: one model for the response times in the critical region (i.e., *written*) and a model on the accuracy in making the correct choice of word at the critical region (i.e., *written*).

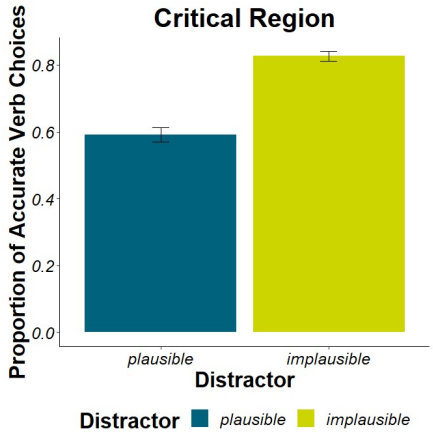


Figure 3.4: *Proportion of correct responses by condition in experiment 8. Error bars represent standard error.*

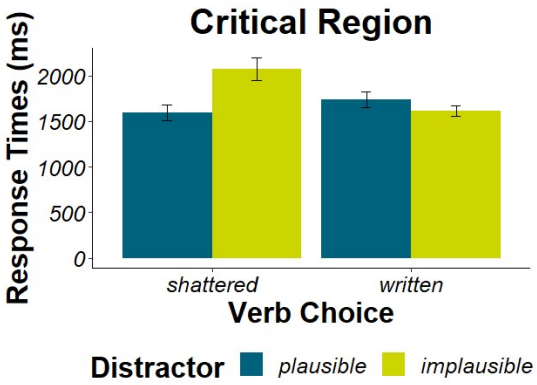


Figure 3.5: *Mean reading times at the critical region in experiment 8. Error bars represent the standard error of the means. The word written is the accurate verb choice whereas the verb shattered is the inaccurate one.*

Condition	Critical region
Implausible Distractor / Accurate Verb	1612.05 (59.82)
Implausible Distractor / Inaccurate Verb	2072.46 (122.54)
Plausible Distractor / Accurate Verb	1735.96 (85.27)
Plausible Distractor / Inaccurate Verb	1594.24 (87.61)

Table 3.8: *Mean response times (ms) per condition and accuracy in the critical region in experiment 8. The SE is given in parenthesis.*

Response Times

Average raw response times by condition in the final critical region are given in table (3.8) and plotted in figure (3.5). The standard error is given in the parentheses. The mean of the posterior distributions over the experimental fixed effect parameters in the critical region, along with the 95% highest posterior density interval (HPDI) are summarised in table (5.16, p.141), which show where the most plausible parameter values for these fixed effect parameters lie based on the given data. The posterior distributions for the critical region are plotted in figure (5.19, p.148).

The model for the response times in the critical region provided evidence for an effect of Distractor Plausibility in the region. The posterior distribution for the effect of Distractor Plausibility showed a credible interval with positive values only ($\hat{\beta}$=

0.05, 95% CrI = [0.03, 0.07], P($\hat{\beta}$ > 0) = 1), suggesting overall shorter response times in the critical region for the sentences containing a plausible distractor as opposed to sentences containing an implausible distractor noun. The posterior distribution for the interaction between Distractor Plausibility and Verb Choice also showed a positive credible interval ($\hat{\beta}$= 0.07, 95% CrI = [0.05, 0.09], P($\hat{\beta}$ < 0) = 1), suggesting that response times for the correct verb in the region were faster when the distractor was implausible. The posterior distribution for the fixed effect of Verb Choice was not informative as it did not provide any evidence for an effect in the response times with the credible interval for the effect spanning over negative and positive values ($\hat{\beta}$= 0.02, 95% CrI = [-0.01, 0.06], P($\hat{\beta}$ < 0) = 0.80).

Accuracy

Figure (3.4) summarises the mean proportion of correct responses in each condition. The mean of the posterior distribution over the fixed effect and the 95% CI in the accuracy model are summarised in table (5.15, p.140). The posterior distribution of the effect for the accuracy model is plotted in figure (5.20, p.148).

The posterior distribution for the fixed effect in the accuracy model provided evidence for a very strong effect of Distractor Plausibility. The credible interval has a range of positive values only, with an estimate of ($\hat{\beta}$= 0.71, 95% CrI = [0.57, 0.86], P($\hat{\beta}$ > 0) = 1), suggesting that subjects were more accurate in selecting the correct verb in the critical region when the sentences contained a distractor that was implausible rather than when the distractor was plausible.

3.4.7 Discussion

The analysed results of this study were quite compatible with what was expected but they also allowed for the observation of a particularly fascinating find.

There were indeed attraction effects in the response times in the critical region with participants overall being faster in their choosing a verb regardless of which one they actually chose, when they were presented with the plausible distractor word. More so, participants were faster in their choosing the correct continuation verb for the sentence when the distractor in the sentence was implausible. Which would suggest that when not encountering a potentially matching distractor word they did not really entertain any other option for a verb choice, while that was not the case when a matching distractor was presented; something which appeared to interact with their choosing the correct verb.

What was of great interest here, was that the plausibility of the distractor word appears to have an effect in how accurate participants were in choosing correctly. Importantly, when presented with a plausible distractor word, the proportion of ac-

curate choices fell by a lot; suggesting that when encountering a word later in the sentence that might fit with what was to follow (so when making a choice for the last word of the sentence in this case), participants were quite split in their choices. The proportion of accurate responses when presented with a plausible distractor is around 60%, suggesting that participants were torn in their choosing a continuation word for the sentence (as seen in figure 3.4). That is while knowing that it was the end of the sentence that they were choosing a word for, as in the presentation of the two options a full stop was also included. Furthermore, they were being given the option of a word that would allow for an overall grammatical sentence (the word *shattered*). Nevertheless, participants still made the inaccurate choice of word in around 40% of the cases. This then shows that it was the plausibility of the intervening distractor word that appears to confuse participants in choosing the correct word.

That finding then could be characterised as an effect of semantic attraction between an intervening distractor word and the choice of verb in the end, but it is not necessarily and straightforwardly compatible with what cue-based models would predict, or at least some versions of them which suggest attraction being a mainly error-driven process (Lago et al., 2015; Wagers et al., 2009), due to the asymmetry between the evidence for attraction effects in ungrammatical (i.e., Hammerly et al., 2019; Patson and Husband, 2016; Wagers et al., 2009) versus the until recently scarce findings in grammatical sentences (i.e., Häussler, 2009; Laurinavichyute and Von der Malsburg, 2024; Nicol et al., 1997; Pearlmutter et al., 1999). If attraction effects arise as an attempt to rescue an otherwise ungrammatical sentence, then those effects should not appear at all in grammatical sentences, and even more so, when given the option of grammaticality. The most important finding here is that in this task, attraction effects were present in all sentences, as the existence of a distractor noun word could determine (to a rather significant portion based on the findings of this task) the choices people make when building sentences. Being presented with plausible distractor words apparently confused participants in whether to go for a grammatical verb which would allow for making meaning for the sentence overall, or for a verb that might match a word they have encountered but nonetheless makes the sentence ungrammatical.

3.5 Experiment 9: Rapid Serial Visual Presentation Task I

The results collected in the previous experiment (experiment 8) had a quite unexpected turn. To exclude the possibility that the surprising findings were an artefact of the task of choice (maze task) used for those purposes, another task was chosen as a control. The task of choice was an RSVP judgement study, where participants were asked to judge the plausibility of the presented sentences in a controlled timed manner (Potter, 1984). As there have been studies that report attraction effects in offline measures too (Nicol et al., 1997, Wagers et al., 2009), if the previous attraction results were not simply induced by the maze task, they should appear in this following RSVP task as well. The same experimental sentences were run in an RSVP acceptability judgement study then to check whether the previously found results still hold [1].

3.5.1 Participants

72 native British English participants were recruited for this task (32 females - 40 males, with a median age of 32.5). The recruitment was via the Prolific Academic platform (www.prolific.co). Participants had self-reported to have no hearing or literacy impairment, normal or corrected to normal vision and no language disorder. They were asked for informed consent before the completion of the experiment and were compensated for their participation after finishing the task (£4.50).

3.5.2 Materials

The experimental materials from experiment 8 were used in a 2x2 design, for a total of 40 experimental sentences. For each of the items, two of the four conditions involved globally implausible sentences whereas the other two were plausible. The plausible sentences were identical to the ones used in experiment 8 (an example of experimental sentences is given in table 3.7). The implausible sentences were the sentences from experiment 8 modified, in such a way so that what used to be the foil in that experiment (i.e., the verb *shattered* in experimental sentences in table 3.7) here came to be the relative clause verb (i.e., the verb *shattered* in the example experimental sentence in table 3.9). Importantly, the manipulation of the plausibility of the distractor is characterised on the basis of the relationship between the distractor noun (i.e., *cup/tie*) and the verb *shattered*, as the *cup* is a plausible argument of *shattered* by

[1] The data collection for experiment 9 (Rapid Serial Visual Presentation Task I) was run by the MA student Athirah Mursyidah Nor Zailan whose help I greatly appreciate.

being *shatterable*, whereas that is not the case for the *tie* with it being an implausible argument. Both of these distractor words are actually implausible arguments of *written*, so in the plausible sentence conditions in table 3.9 the plausible and implausible distractor manipulation are relative to the verb *shatter* and not to *write*.

Sentence	Distractor	Sentence
plausible	*plausible*	Sue remembered the letter that the butler with the **cup** had recently **written**.
	implausible	Sue remembered the letter that the butler with the **tie** had recently **written**.
implausible	*plausible*	Sue remembered the letter that the butler with the **cup** had recently **shattered**.
	implausible	Sue remembered the letter that the butler with the **tie** had recently **shattered**.

Table 3.9: Example of experimental item in experiment 9

As fillers were used 80 sentences with different structures and an equal number of plausible and implausible. Great care was taken so that the word that would deem them plausible or not would be close to the end of the sentence, to match the structure of the experimental items (as the word that defines their overall sentence plausibility is the word at the end i.e., *written* for plausible and *shattered* for implausible in the example experimental items in table 3.9). Filler sentences were also controlled for length, so that they would approximately match the length of the experimental items. Overall the tested sentences were equally balanced in the number of plausible vs. implausible (60 + 60 of each kind).

All the experimental stimuli for the experiment can be found on p.193.

3.5.3 Procedure

The PC Ibex software was used to code this experiment (Drummond, 2013; Zehr & Schwarz, 2018a; Zehr & Schwarz, 2018b), while it was run with participants recruited via the Prolific Academic platform (www.prolific.co). Four counterbalanced lists were created so that participants encountered one condition of each experimental sentence only once. In order to encourage high levels of attention to the task, the settings were set such that the task could be run only on a desktop device. Participants were asked to read the sentences that would appear word by word in front of them and then make a judgement on the sentence's plausibility. The words in each sentence would appear one by one automatically and participants were informed that they did not need to do anything for words to appear. Each sentence was followed by the same question of whether the sentence was plausible or not, and participants could either click on the screen in front of them "yes" or "no", or they could press the *Y* key for

yes or the *N* key for no on their keyboard. For each new trial sentence to begin, participants were asked to click on their screen on a prompt saying **Next Sentence**. Each session would begin with a practice session to familiarise the participants with the task, which involved three sentences that they were asked to judge in terms of each sentence's plausibility, with that plausibility being very easily discerned. The task lasted approximately 30 minutes.

3.5.4 Analysis

Before analysing the data, response times for participants' judgement below 100 ms were removed as they were considered too extreme. Three Bayesian linear mixed effects models were fitted to the data with the factors Sentence Plausibility, Distractor Plausibility and their interaction (Sentence Plausibility $\times$ Distractor Plausibility) treated as fixed effects, while Subjects and Items were treated as random effects in the models. Sentence and Distractor Plausibility had two levels each, with Sentence Plausibility coded as -1 (plausible) and 1 (implausible), while Distractor Plausibility was coded as 1 (plausible) and -1 (implausible). The statistical parameters for the statistical models were the same with the ones set for experiment 3. For the model reported on response times here, accuracy has not been considered as a factor.

3.5.5 Results

Three models will be reported and discussed for this task : one model for the response times to the task, one model for the participants' accuracy in their responses and one model on the acceptance rates of the sentences.

Response Times

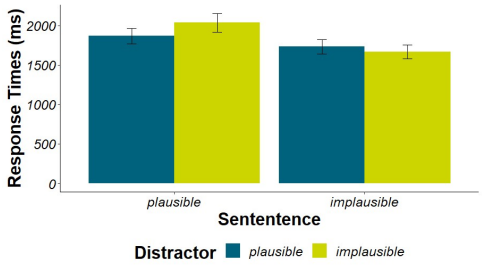


Figure 3.6: *Mean response times in experiment 9. Error bars represent SE.*

Condition	Response Times (ms)
Plausible S. / Plausible D.	1868.18 (96.78)
Plausible S. / Implausible D.	2037.44 (118.12)
Implausible S. / Plausible D.	1731.96 (93.19)
Implausible S. / Implausible D.	1666.24 (87.35)

Table 3.10: *Mean response times (ms) per condition in experiment 9. The SE is given in parenthesis.*

Average raw response times by condition are summarised in table (3.10) and plotted in figure (3.6). The mean of the posterior distribution over the fixed effect parameters for the response times, along with the 95% highest posterior density interval

are summarised in table (5.18, p.141), while the posterior distributions are plotted in figure (5.21, p.148).

The analysis of the response times in the task provided evidence for an effect of Sentence Plausibility, as the credible interval spanned over negative values only, with the Sentence Plausibility estimate being ($\hat{\beta}$= -0.04, 95% CrI = [-0.07, -0.02], P($\hat{\beta}$ < 0) = 0.99), suggesting overall faster response times for the implausible compared to the response times for the plausible sentences. The posterior distribution for the other fixed effects however, proved to be uninformative as the credible intervals for Distractor Plausibility as well as for the interaction between Sentence Plausibility and Distractor Plausibility spanned across 0, suggesting no evidence for an effect. More specifically, the estimate for the Distractor Plausibility was ($\hat{\beta}$= -0.02, 95% CrI = [-0.06, 0.02], P($\hat{\beta}$ < 0) = 0.80) while the estimate for the interaction of Sentence Plausibility and Distractor Plausibility was ($\hat{\beta}$= 0.02, 95% CrI = [-0.02, 0.05], P($\hat{\beta}$ > 0) = 0.84).

Accuracy and Acceptance Rates

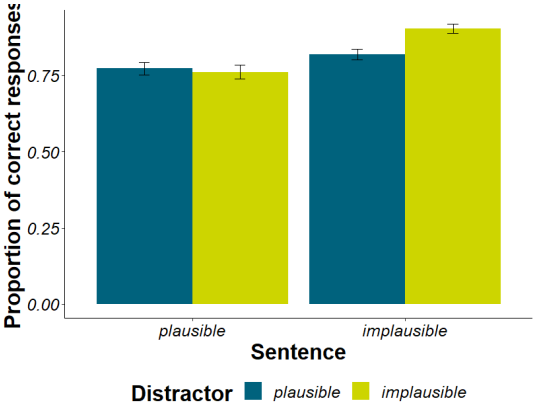


Figure 3.7: *Mean accuracy in experiment 9. Error bars represent the standard error of the means.*

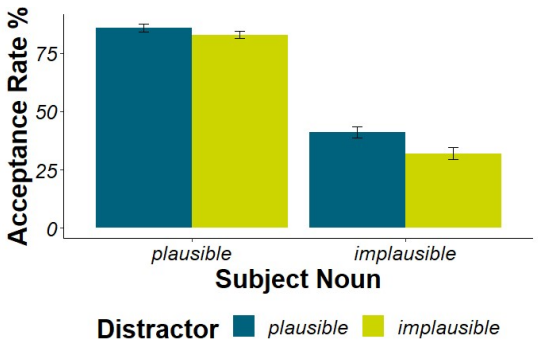


Figure 3.8: *Mean acceptance rates in experiment 9. Error bars represent the standard error of the means.*

The mean proportion of correct responses in each condition are summarised in figure (3.7) and the mean acceptance rates are shown in figure (3.8). Table (5.17, p.141) summarises the mean of the posterior distribution over the experimental fixed effect parameters and the 95% CI in the accuracy model, while table (5.19, p.141) summarises the same for the acceptability rates model. The relevant posterior distributions for these two models are plotted in figure (5.22, p.148) for the accuracy model and in figure (5.23, p.148) for the acceptability rates model.

The analysis of the accuracy as well as the acceptance rate data showed that the posterior distributions of the relevant models gave rise to evidence for effects in both of them. In particular, for the accuracy model both the Sentence Plausibility and the interaction between Sentence Plausibility and Distractor Plausibility showed strong

effects, while the estimate of the Distractor Plausibility was also in the direction of an effect despite spanning over 0 ($\hat{\beta}$= -0.23, 95% CrI = [-0.49, 0.01], P($\hat{\beta}$ < 0) = 0.96). The estimate of the Sentence Plausibility provided very strong evidence for the effect spanning over positive values only ($\hat{\beta}$= 0.42, 95% CrI = [0.28, 0.56], P($\hat{\beta}$ > 0) = 1), suggesting that overall participants were more accurate in the implausible sentences, while the estimate for the interaction spanned over negative values only ($\hat{\beta}$= -0.25, 95% CrI = [-0.38, -0.13], P($\hat{\beta}$ < 0) = 1), indicating that a plausible distractor in a plausible sentence led to a higher proportion of accurate responses compared to an implausible distractor in the same type of sentence whereas a plausible distractor in an implausible sentence led to a lower proportion of correct responses compared to an implausible distractor in the same sentence.

The posterior distribution of the fixed effects in the acceptance rate model provided very strong evidence for all three effects. The credible interval for Sentence Plausibility spanned over negative values only ($\hat{\beta}$= -1.56, 95% CrI = [-2.91, -0.14], P($\hat{\beta}$ < 0) = 0.98), suggesting that sentences were more likely to be accepted when plausible than when implausible. The estimates for the Distractor Plausibility and the interaction between Sentence Plausibility and Distractor Plausibility spanned over positive values only, with the estimate for the Distractor Plausibility being ($\hat{\beta}$= 9.09, 95% CrI = [7.13, 11.07], P($\hat{\beta}$ > 0) = 1) indicating that a plausible distractor noun led to overall higher acceptance rates and the estimate for the interaction of Sentence Plausibility and Distractor Plausibility being ($\hat{\beta}$= 3.23, 95% CrI = [1.72, 4.59], P($\hat{\beta}$ > 0) = 1), signifying that a plausible distractor in the implausible sentences led to higher acceptance rates compared to an implausible distractor, as well as higher acceptance rates compared to the same distractor in plausible sentences.

3.5.6 Discussion

The results of this task are compatible with previous findings for this type of sentences.

It is evident in the response times that participants are much faster to detect and judge an implausible sentence rather than a plausible one. When it comes to just judging the plausibility of a sentence, participants appear not to focus on the distractor word at all, as there was no such effect in the results for the response times. That finding is quite compatible with the finding in Cunnings and Sturt (2018), where early eye-tracking measures (first pass fixations, regressions) provided evidence that participants detected the plausibility or implausibility of the tested sentences, while the plausibility of the distractor word had no effect at that stage.

Accuracy data also followed this pattern, as when it comes to how accurate participants were in their judgements of the tested sentences, implausibility also led them to

higher accuracy compared to the plausible sentences overall. Accuracy data allowed for observing attraction effects in the tested sentences too, as plausible distractors in implausible sentences led to more inaccurate responses, suggesting that the presentation of a distractor that is a plausible complement of a verb followed shortly by such a verb led participants to judge globally implausible sentences as plausible.

The acceptance rate data were quite similar. The plausibility of the sentence highly determined if a sentence would be accepted or not. Attraction effects were also observed here, as a plausible distractor would deem a sentence as more plausible than the inclusion of an implausible distractor. A potential match between a distractor word and a verb led participants to accept the implausible sentences as plausible overall, despite them being globally implausible.

When it comes to comparing the data from this experiment to the ones from the preceding maze task (experiment 8), it is not an exact match. The data here are quite compatible with previous research in attraction effects, where it is clear that participants very quickly and accurately first determine whether a sentence is plausible or not, while there are also signs of attraction between an intervening matching distractor word and a verb in implausible sentences. However, there was no sign of a significant difference in the results when it comes to globally plausible sentences being affected by the plausibility of such a distractor neither in the speed, nor in accuracy or their acceptance. That could indeed be attributed to the task itself, as no such effects have been found in previous research, so they might be due to the use of the maze task, but it is definitely not conclusive through this study.

3.6 Experiment 10a: Plausibility Study I for Maze Task IV

The next study (experiment 11) attempted to investigate whether the complexity of the sentences that were used so far across these previous experiments had any effect on the attraction phenomena observed. Admittedly, the sentences used as experimental items have been rather structurally complex and this added difficulty might have resulted in amplifying the attraction effects that were found both in Cunnings and Sturt (2018), experiment 7 and experiment 8. To do so, new experimental sentences were to be built that were in part similar to the ones used in the previous experiments but did not have such a complex structure; a structure that involved the verb of interest buried in an object relative clause, as well as distractor nouns that are enclosed in prepositional phrases in the same long relative clause (example of experimental items are in tables 3.5 for experiment 7 and in table 3.7 for experiment 8). The goal for the next experiment (experiment 11) was to simplify the sentences that would act as experimental items by making two important changes in the syntactic structure that was followed in the previous experimental items. First, the existence of a relative clause was naturally retained for consistency with previous items and it also included the nouns that acted as the distractors, however those distractor nouns were not part of a prepositional phrase anymore, but they were syntactic objects within this relative clause. At the same time, the verb of interest where the attraction effects would be sought out to be observed, became part of the main sentence by being the main verb of the experimental items.

In experiment 10a then the goal was to test the plausibility of potential experimental sentences with a simpler structure and implementing the changes discussed above (in a simple plausibility study) for them to be used in experiment 11.

3.6.1 Participants

24 native British English participants were recruited for this plausibility study (18 females - 6 males, median age of 34.7). The recruitment was carried out via the Prolific Academic platform (www.prolific.co), where all participants during their prescreening had reported to have normal or corrected to normal vision, no hearing impairment and no literacy or language disorder. Informed consent was provided before running the experiment and all participants were compensated for their participation after the end of the study (£0.63).

3.6.2 Materials

The sentences for this task were 40 simple active sentences that involved a noun in the subject position (i.e., as *author / journalist / cat / rain* in the example experimental item in table 3.11) followed by a verb in past perfect (i.e., *had...written*), an adverb (i.e., *carefully*), and another noun phrase that also included an adjective (i.e., *a short novel*). Each item had four conditions, where two of them were expected to be judged as equally plausible (conditions 1 + 2 as seen in table 3.11) whereas the other two as implausible (conditions 3 + 4 as seen in table 3.11).

C	Sentence
1	The author had carefully written a short novel.
2	The journalist had carefully written a short novel.
3	The cat had carefully written a short novel.
4	The rain had carefully written a short novel.

Table 3.11: Example of experimental item in experiment 10a

As the 40 experimental sentences included an equal number of plausible and implausible conditions, only a set of 10 fillers was used as control, with 5 of those sentences being highly plausible and 5 being highly implausible.

3.6.3 Procedure

The procedure followed in this task was identical with the one followed in experiment 6.

3.6.4 Analysis & Results

The analysis followed the analysis of the data in experiment 6. The collected data were analysed in R (Team R Core, 2020), with each experimental sentence having four conditions for which individual ratings per item were calculated, together with an average mean for each condition. The mean by condition is given in table 3.12 :

Condition	Mean
S-1	6.09
S-2	5.65
S-3	2.63
S-4	2.49

Table 3.12: By condition mean ratings in experiment 10a

The mean ratings of the 10 fillers were used to check that participants paid attention and gave accurate judgements. The analysis of the mean ratings for the fillers

confirmed that the quality of the data was acceptable. Despite the generally good mean rating on the by condition average, on the item by item basis some problems were detected, with plausibility ratings for some of the conditions of each item not being discerning enough for the purposes of experiment 11.

3.7 Experiment 10b: Plausibility Study II for Maze Task IV

As some of the sentences that were built and thereupon tested in experiment 10a appeared to not have the characteristics that were required for the subsequent planned experiment (Experiment 11), another plausibility study was run with the revision of some of the problematic sentences of experiment 10a as well as new additions in order to reach the target of 40 sentences that was deemed as an adequate number of experimental items for experiment 11.

3.7.1 Participants

An additional 16 participants were recruited for the revised or newly formed sentences via the same platform (www.prolific.co) (median age of 36.8, 8 females - 4 males). As in the previous task, they were self-reportedly native British English speakers, without hearing, visual, literacy or language impairments. Informed consent was required for participation in the experiment and compensation for subjects' participation in the study was given after the completion of the task (£0.45).

3.7.2 Materials

The materials included 14 sentences from experiment 10a, where the judgement of the plausibility was not particularly distinct and did not reach the levels desired for the purposes of experiment 11 between the four conditions. 14 sentences were slightly more polished to achieve distinct plausibility differences between the conditions. 10 filler sentences were used as control of plausibility, with half being very plausible and expected to be on the scale of 6-7 and the other half being very implausible with the expectation that they are judged on the scale 1-2.

3.7.3 Procedure

The procedure followed was the same as the one in experiment 6.

3.7.4 Analysis & Results

The collected data were analysed in similar fashion to the data from experiments 6 and 10a in R (Team R Core, 2020). Each item had four conditions for which a by item as well as a mean average were calculated as shown below (3.13):

Condition	Mean
S-1	6.41
S-2	6.17
S-3	2.44
S-4	2.23

Table 3.13: By condition mean ratings in experiment 10b

The analysis of the fillers confirmed that the ratings were indeed given with attention from the selected participants. Looking on the individual by item and by condition rating allowed for seeing a clear distinction in judgement of plausibility between conditions that would be essential for the next experiment.

3.7.5 Discussion

Testing those extra experimental items allowed for creating a set of 40 sentences for experiment 11 that were all well formed and with very clear plausibility conditions, as in conditions 1 and 2 being very highly and equally plausible and conditions 3 and 4 being highly and equally implausible.

3.8 Experiment 11: Maze Task IV

This task sought to investigate whether the previously found attraction effects, both in the preceding experiments, as well as in the original Cunnings and Sturt (2018) could be a result of the structural complexity of those tested sentences. Therefore the purpose of this study was to test whether attraction effects lead to grammaticality illusions in less complex sentences than those used in the previous studies.

3.8.1 Participants

96 British English natives were recruited via the Prolific Academic platform (www.prolific.co) to participate in this study (participants' median age: 35, 60 females - 36 males). At the pre-screening stage before their participation, all participants had self-reported to have normal or corrected to normal vision, no literacy or hearing impairments and no language difficulties. Informed consent was asked to be provided by all participants before participating in the study and compensation was given after the completion of the task (£6).

3.8.2 Materials

The sentences built for the purposes of experiment 11 needed to be less complex than the ones in the previous experiments. To achieve that, the verb in question was transferred to the main sentence rather than it being in a relative clause (i.e., the verb *interviewed* in the example sentence in table 3.14). The structure of the newly formed sentences involved a noun (i.e., *officer / dog*) followed by a relative clause which included a verb and another noun (i.e., *that liked the reporter / walk*), an adverb (i.e., *immediately*), the verb of interest (i.e., *interviewed*), another noun phrase as a complement to the verb (i.e., *the witnesses*) and a prepositional phrase (i.e., *at the scene*). To increase the distance between the intervening noun of the relative clause and to keep consistent with the tested experimental items in the previous experiments, the past perfect tense was used and an adverb was also preceding the verb. As a result, between the potentially plausible complement of the verb and the verb itself there were always two intervening words (i.e., *reporter / walk* ***had immediately*** *interviewed* in the example sentence in table 3.14).

All the experimental stimuli for the experiment can be found on p.200.

3.8.3 Procedure

The experiment was coded with the PC Ibex Farm software (Drummond, 2013; Zehr & Schwarz, 2018a; Zehr & Schwarz, 2018b) and was then run via the Prolific Aca-

Noun	Distractor	Sentence
plausible	*plausible*	The **officer** that liked the **reporter** had immediately interviewed the witnesses at the scene.
	implausible	The **officer** that liked the **walk** had immediately interviewed the witnesses at the scene.
implausible	*plausible*	The **dog** that liked the **reporter** had immediately interviewed the witnesses at the scene.
	implausible	The **dog** that liked the **walk** had immediately interviewed the witnesses at the scene.

Table 3.14: Example of experimental item in experiment 11

demic platform which helped with the recruitment of participants (www.prolific.co). Four counterbalanced lists were created, to ensure that the stimuli were encountered only once during the course of the experiment. Each list contained only one condition of each stimulus sentence. The task was similar in design to the one described in detail for experiment 4, however no comprehension questions manipulation was included for the present task. The experiment lasted approximately for 40 minutes.

3.8.4 Analysis

Extreme response times (above 10.000 ms) were removed from the data before the analysis, while trials were only included if participants were correct in their selection of words up in all positions preceding the critical region.

Three Bayesian linear mixed effects models were fitted to the data including the factors Subject Noun Plausibility, Distractor Plausibility as well as their interaction (Subject Noun Plausibility $\times$ Distractor Plausibility) treated as fixed effects in the models, while Subjects and Items as random effects. Each of the fixed effects factors (Subject Noun Plausibility and Distractor Plausibility) had two levels; Subject Noun Plausibility was coded as 1 (plausible) and -1 (implausible) and Distractor Plausibility was coded as -1 (plausible) and 1 (implausible).

The statistical models had the same statistical parameters with the ones in experiment 3.

3.8.5 Regions of Interest

For this task two regions were analysed and will be reported here: the critical and the spillover regions. The critical region included the verb in the relative clause in the stimuli sentences (i.e., *interviewed*), while the spillover region included the word following that verb (i.e., *the*).

3.8.6 Results

The three models reported for this experiment are on the accuracy in selecting the correct word in the critical region and on the response times for the two regions of interest : the critical region (i.e., *interviewed*) and the spillover region (i.e., *the*).

Response Times

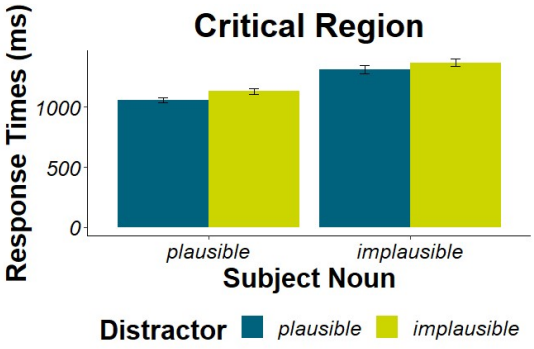


Figure 3.9: *Mean response times at the critical region in experiment 11. Error bars represent standard error.*

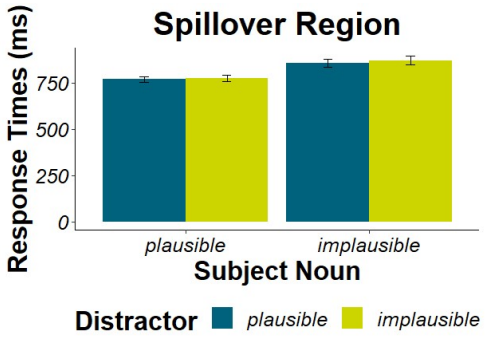


Figure 3.10: *Mean response times at the spillover region in experiment 11. Error bars represent the standard error of the means.*

Condition	Critical region	Spillover region
Implausible Noun / Implausible Distractor	1370.42 (31.83)	872.03 (23.13)
Implausible Noun / Plausible Distractor	1311.66 (36.01)	857.37 (23.14)
Plausible Noun / Implausible Distractor	1130.68 (25.46)	773.60 (16.53)
Plausible Noun / Plausible Distractor	1057.10 (22.64)	768.82 (14.42)

Table 3.15: *Mean response times (ms) per condition by region in experiment 11. The SE is given in parenthesis.*

Table (3.15) summarises the average raw response times by condition in the critical and spillover regions whereas these response times are plotted in figures (3.9 & 3.10) for each region of interest. The standard error is given in the parentheses. Table (5.21, p.142) summarises the mean of the posterior distributions over the experimental fixed effect parameters in the the two regions of interest, the critical and the spillover regions, along with the 95% highest posterior density interval (HPDI). These Bayesian credible intervals indicate where the most plausible parameter values for these fixed effects parameters lie, given the data. The posterior distributions are plotted in figure (5.24, p.149) and in figure (5.25, p.149) for the critical and the spillover regions respectively.

After analysing the response times in the critical region, the posterior distributions showed that there was evidence for an effect of Noun Plausibility as well as an effect of Distractor Plausibility. The credible intervals for both spanned over either negative

or positive values only, with the Noun Plausibility estimate being ($\hat{\beta}$= -0.09, 95% CrI = [-0.12, -0.06], P($\hat{\beta}$ < 0) = 1), suggesting that overall participants were slower when the subject noun was implausible, while the Distractor Plausibility estimate was ($\hat{\beta}$= 0.03, 95% CrI = [0.01, 0.04], P($\hat{\beta}$ > 0) = 1), implying faster response times in the region for sentences containing a plausible distractor compared to sentences with an implausible distractor. The posterior distribution for the interaction of these two effects (Noun Plausibility and Distractor Plausibility) was uninformative, with an estimate of ($\hat{\beta}$= -0.00, 95% CrI = [-0.01, 0.01], P($\hat{\beta}$ > 0) = 0.56).

The posterior distributions for the response times in the spillover region showed an effect of Noun Plausibility with the credible interval for the Noun Plausibility spanning over positive values only ($\hat{\beta}$= 0.04, 95% CrI = [0.03, 0.06], P($\hat{\beta}$ > 0) = 1), suggesting that participants were overall slower in their response times when the subject noun was implausible. The posterior distributions for the Distractor Plausibility effect and the interaction between Noun Plausibility and Distractor Plausibility showed no evidence for effects in the region, with the estimate for the Distractor Plausibility being ($\hat{\beta}$= -0.01, 95% CrI = [-0.02, 0.01], P($\hat{\beta}$ < 0) = 0.88), whereas that of the interaction of Noun Plausibility and Distractor Plausibility being ($\hat{\beta}$= -0.00, 95% CrI = [-0.01, 0.02], P($\hat{\beta}$ < 0) = 0.52).

Accuracy

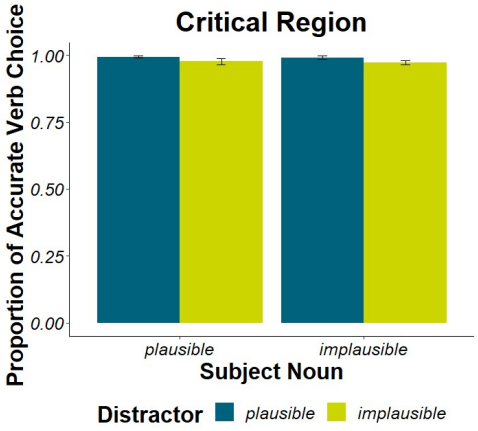


Figure 3.11: *Proportion of correct responses by condition in experiment 11. Error bars represent standard error.*

The mean proportion of correct responses in each condition is summarised in figure (3.11). The mean of the posterior distribution over the fixed effects and the 95% CI in the accuracy model are summarised in table (5.20, p.141) and the posterior distribution of the effects for the accuracy model is plotted in figure (5.26, p.150).

The analysis of the accuracy in selecting the correct verb data showed that the pos-

terior distribution for the fixed effects in the accuracy model provided evidence for a strong effect of Distractor Plausibility. The credible interval has an estimate of ($\hat{\beta}$= 0.69, 95% CrI = [0.02, 1.43], P($\hat{\beta}$ > 0) = 0.97), suggesting that subjects were more accurate in selecting the correct verb in the critical region when the sentences contained a distractor that was plausible rather than when the distractor was implausible. The posterior distribution provided uninformative evidence for the effect of Noun Plausibility and the interaction between Noun Plausibility and Distractor Plausibility, as their credible intervals spanned over positive and negative values alike. The estimate for the Noun Plausibility was ($\hat{\beta}$= -0.26, 95% CrI = [-0.95, 0.42], P($\hat{\beta}$ < 0) = 0.78) while the estimate for the interaction Noun Plausibility and Distractor Plausibility was ($\hat{\beta}$= 0.26, 95% CrI = [-0.39, 0.97], P($\hat{\beta}$ > 0) = 0.77).

3.8.7 Discussion

The results of this task proved to be quite informative. As expected, the response times data showed that overall in sentences that contained a plausible subject noun participants had a faster response in choosing a word both in the critical as well as in the spillover positions. Importantly, those also showed evidence for attraction effects, with participants making choices for the critical region faster when there was a plausible distractor word in the sentence compared to when the distractor was implausible. This facilitation in the speed of making a choice when it comes to choosing a verb is compatible with the previous finding in experiment 7 and it is both for the globally plausible as well as for the globally implausible sentences. This effect was found in the critical region only, as by the time participants had to make a choice for the spillover region the plausibility of the distractor noun was not significant.

A similar effect was found in the accuracy data, as a plausible distractor noun led to higher levels in accuracy both in the plausible as well as in the implausible sentences too. Surprisingly, the plausibility of the subject noun appeared not to play a role in the accuracy in choosing the correct word for the critical region.

3.9 Experiment 12: Rapid Serial Visual Presentation Task II

Similar to experiment 9 which was run as a control experiment for the experimental items used in experiment 8, experiment 12 was run to rule out once again potential maze task induced effects in the previous findings of experiment 11. The newly formed experimental sentences from experiment 11 (i.e., an example of what the experimental items looked like can be found in table 3.14) were run in an acceptability RSVP task [2].

3.9.1 Participants

72 British English natives were recruited via the Prolific Academic platform (www.prolific.co) to complete this task (32 females - 40 males, median age of sample: 32.5). At the pre-screening, participants had reported to have neither any visual or hearing impairment nor any literacy or language disorders. Informed consent was asked before participation in the task and compensation was given after the completion of the experiment (£4.50).

3.9.2 Materials

The experimental sentences were identical with the ones in experiment 11. Items were in a 2×2 design, crossing the plausibility of the subject noun (i.e., *officer / dog*) and that of the distractor noun (i.e., *reporter / walk*), on the basis of how much of a plausible argument they could be for the main verb of the sentence (i.e., *interviewed* as seen in the example experimental item in table 3.16).

Subject Noun	Distractor	Sentence
plausible	*plausible*	The **officer** that liked the **reporter** had immediately interviewed the witnesses at the scene.
	implausible	The **officer** that liked the **walk** had immediately interviewed the witnesses at the scene.
implausible	*plausible*	The **dog** that liked the **reporter** had immediately interviewed the witnesses at the scene.
	implausible	The **dog** that liked the **walk** had immediately interviewed the witnesses at the scene.

Table 3.16: Example of experimental item in experiment 12

All the experimental stimuli for the experiment can be found on p.210.

[2]The data collection for experiment 12 (Rapid Serial Visual Presentation Task II) was run once again by the MA student Athirah Mursyidah Nor Zailan whose help I greatly appreciate.

3.9.3 Procedure

To code this experiment PC Ibex software was used (Drummond, 2013; Zehr & Schwarz, 2018a; Zehr & Schwarz, 2018b), while the Prolific Academic platform (www.prolific.co) was used to run it by recruiting participants. To ensure that each participants encounters one condition of each experimental sentence only once in each session, four counterbalanced lists were created. To ensure high levels in participants' attention, the task could run only on a desktop device and not on phones or tablets. The design of the task was similar with the that of experiment 9. The task lasted approximately 30 minutes.

3.9.4 Analysis

Response times below 100 ms were removed as they were considered too extreme before analysing the data. During the analysis, three Bayesian linear mixed effects models were fitted to the data with the factors Subject Noun Plausibility, Distractor Plausibility and their interaction (Subject Noun Plausibility $\times$ Distractor Plausibility) treated as fixed effects, while Subjects and Items were treated as random effects in the models. Subject Noun Plausibility and Distractor Plausibility had two levels each, with Subject Noun Plausibility coded as 1 (plausible) and -1 (implausible), while Distractor Plausibility was coded as -1 (plausible) and 1 (implausible). The rest of the analysis followed the analysis of experiment 9, while the statistical parameters for the statistical models were the same with the ones set for experiment 3. For the model reported on response times here, accuracy has not been considered as a factor.

3.9.5 Results

The three models reported and discussed for this task include one model for the response times to the task, one model for the participants' accuracy in their responses and one model on the acceptance rates of the sentences.

Response Times

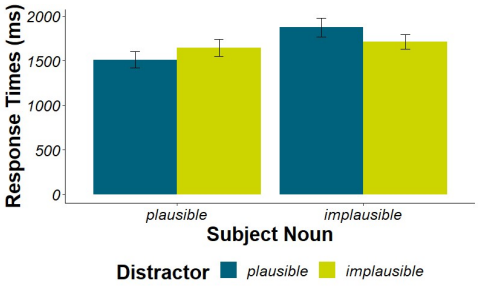


Figure 3.12: *Mean response times in experiment 12. Error bars represent SE.*

Condition	Response Times (ms)
Plausible N. / Plausible D.	1508.92 (91.31)
Plausible N. / Implausible D.	1645.05 (94.40)
Implausible N. / Plausible D.	1872.74 (104.04)
Implausible N. / Implausible D.	1713.51 (80.97)

Table 3.17: *Mean response times (ms) per condition in experiment 12. The SE is given in parenthesis.*

The mean response times in ms are summarised by condition in table (3.17) and plotted in figure (3.12). Table (5.23, p.142) summarises the mean of the posterior distribution over the fixed effect parameters along with the 95% highest posterior density interval for the response times model and the distribution is plotted in figure (5.21).

The posterior distribution of the fixed effects in the response times model based on the analysis of the response times data showed that there is evidence for some effects. The credible intervals of the Subject Noun and the interaction of Subject Noun and Distractor Plausibility did not cross 0 and spanned over either only positive or only negative values. The estimate of the Subject Noun was ($\hat{\beta}$= -0.06, 95% CrI = [-0.10, -0.02], P($\hat{\beta}$ < 0) = 0.99) denoting that sentences with a plausible subject noun had overall faster response times than sentences with an implausible subject noun. At the same time, the estimate of the interaction of Subject Noun and Distractor Plausibility was ($\hat{\beta}$= 0.03, 95% CrI = [0.01, 0.06], P($\hat{\beta}$ > 0) = 0.99), showing that a plausible distractor led to shorter response times in plausible sentences but longer response times in implausible sentences. The estimate of the Distractor Plausibility was uninformative in this model ($\hat{\beta}$= 0.00, 95% CrI = [-0.02, 0.03], P($\hat{\beta}$ > 0) = 0.59).

Accuracy and Acceptance Rates

Figure (3.13) summarises the mean proportion of correct responses by condition while the mean acceptance rates are shown in figure (3.14). The mean of the posterior distribution over the experimental fixed effect parameters and the 95% CI for the accuracy model are given in table (5.22, p.142), while table (5.24, p.142) summarises the same for the acceptability rates model. The relevant posterior distributions for these two models are plotted in figure (5.28, p.150) and in figure (5.29, p.151) for the accuracy and the acceptability rates models respectively.

The analysis of the posterior distribution of the fixed effects for the accuracy model gave evidence for similar effects with the ones based on the posterior distribution for the response times model. There was strong evidence for a Subject Noun

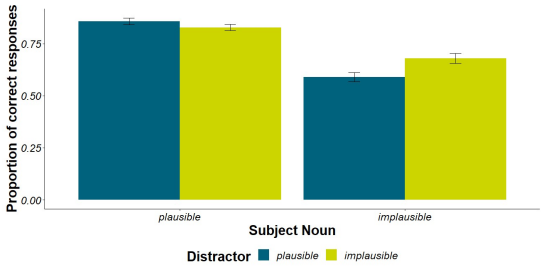


Figure 3.13: *Mean accuracy in experiment 9. Error bars represent the standard error of the means.*

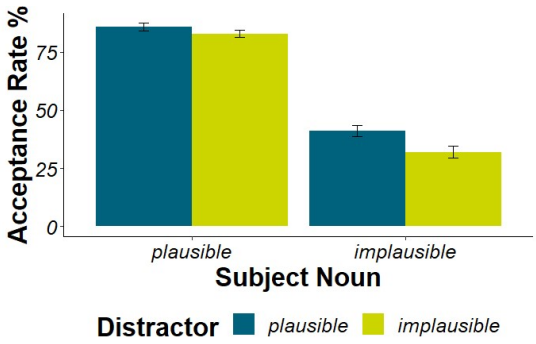


Figure 3.14: *Mean acceptance rates in experiment 9. Error bars represent the standard error of the means.*

effect with an estimate ($\hat{\beta}$= 0.63, 95% CrI = [0.29, 0.96], P($\hat{\beta}$ > 0) = 0.99), suggesting higher levels of accuracy when the subject noun was plausible compared to when implausible and at the same time very strong evidence for an interaction effect between Subject Noun and Distractor Plausibility with an estimate of ($\hat{\beta}$= -0.26, 95% CrI = [-0.45, -0.08], P($\hat{\beta}$ < 0) = 0.99), indicating that a plausible distractor led to higher accuracy in the sentences with a plausible subject noun but a lower accuracy in the sentences with an implausible subject noun. Once again, there was no evidence of a Distractor Plausibility ($\hat{\beta}$= 0.04, 95% CrI = [-0.09, 0.18], P($\hat{\beta}$ > 0) = 0.73).

The posterior distribution of the fixed effects for the acceptability rate model was slightly different, in that there was very strong evidence for a Subject Noun effect only. The credible interval for the Subject Noun effect spanned over positive values only ($\hat{\beta}$= 9.35, 95% CrI = [7.43, 11.22], P($\hat{\beta}$ > 0) = 1), inferring that sentences with a plausible subject noun are considered as more acceptable overall than sentences with an implausible subject noun. The estimates for the Distractor Plausibility ($\hat{\beta}$= -1.39, 95% CrI = [-2.77, 0.04], P($\hat{\beta}$ < 0) = 0.97) and the interaction of Subject Noun and Distractor Plausibility ($\hat{\beta}$= 1.10, 95% CrI = [-0.10, 2.32], P($\hat{\beta}$ > 0) = 0.96) were not informative, however the Distractor Plausibility was in that direction.

3.9.6 Discussion

The results of this study appeared to be rather interesting compared to the ones from previous experiments. The response times data as expected showed that participants were very quick to judge a sentence's plausibility however in a different direction, based on whether the subject noun was itself plausible. That is very dissimilar to the findings in experiment 9 and by extension to the original Cunnings and Sturt (2018), where implausible sentences led to faster response times. The effect here seems to be completely reversed, as there appears to be facilitation in response times when the subject noun was plausible and therefore when the sentence was globally plausible.

At the same time, the effect that the plausibility of the distractor noun had on the response times while quite different compared to the results from experiment 11, was similar to the ones from experiment 9. While in experiment 9 there was only a tendency in that direction, here a plausible distractor noun leads to facilitation in the response times but it leads to delays when the sentence is itself implausible. That is the exact opposite of the attraction effects would appear like and what was seen in experiment 11, as in that case it was observed that a plausible distractor noun led to facilitation in processing rather than leading to a delay.

The accuracy results followed the same pattern. The accuracy in subjects' responses was primarily mediated by the plausibility of the subject noun and by extension the plausibility of the sentence itself, with responses being more accurate when the subject noun was plausible. At the same time the plausibility of the distractor noun led to higher accuracy levels in the plausible sentences but lowered accuracy by a lot in the implausible sentences. That is manifested (albeit numerically) in the acceptability rates as well.

Overall, it appears that when it comes to the timing for making a judgement on a sentence, participants relied heavily on the basic structure of the sentence. They quickly judged whether a sentence was plausible by relying on whether the subject noun and the verb were a good match, and opposite to what happens in most cases with attraction effects, the plausibility of a potentially matching distractor did not lead to any facilitation in the implausible sentences. The plausibility of the distractor was considered in otherwise plausible sentences only, while in implausible sentences it could possibly create more confusion and therefore delays in the timing of the judgement. In combination with the data from the accuracy and the acceptability models, a plausible distractor appears to lead to inaccurate judgements more so for implausible sentences despite the longer response times. That suggests that participants likely entertain the match between the distractor and the verb and eventually deem the whole sentence as more acceptable or plausible than it actually was. That then could be compatible with the idea that the plausibility of a distractor in a sentence could interfere with the processing of that given sentence but incompatible with the notion that such interference works in the direction of facilitated processing (i.e., at least based on the reaction times).

3.10 Effect of Experiment on Distractor Plausibility

After reviewing the findings in all of the above experiments, a comparison model was run to better explore the size of the effect of Distractor Plausibility depending on the individual differences of two of the tasks previously presented. More specifically the model compared the distractor effect in experiment 7 (Maze Task II) and experiment 11 (Maze Task IV), by collapsing their data into a single model. That was done to compare whether the individual differences, namely the much more complex structure in the sentences in experiment 7 and the simpler structure of the sentences in experiment 11 could account for the different findings in those two experiments.

The response times of the two tasks were combined in one data frame, with the same criteria used as in the tasks (i.e. exclusion of extreme values). They were then analysed with a Bayesian linear mixed effects model that assumes a Gaussian distribution, with the effects of Distractor Plausibility and Experiment, and their critical interaction. The levels of Experiment were coded as 1 (for experiment 7: the maze task II) and -1 (for experiment 11: the maze task IV). For the fixed effects, normal priors were set, with a mean value of 0 and the variance of the distribution set to 1.

The analysis of the collapsed model apart from the expected Distractor Plausibility effect, also revealed an effect of Experiment, as well as a critical interaction between Distractor Plausibility and Experiment. The mean of the posterior distributions over the fixed effect parameters together with the 95% highest posterior density interval (HPDI) are given in table 5.25, p.142. The posterior distributions are plotted in figure 5.30, p.151. The posterior distribution for the effect of Experiment showed a relatively large and very clear mean ($\hat{\beta}$= -0.13, 95% CrI = [-0.16, -0.09], P($\hat{\beta}$ < 0) = 1), suggesting that reaction times in experiment 11 were overall longer than those in experiment 7. The posterior distribution for the expected effect of Distractor Plausibility showed a positive credible interval ($\hat{\beta}$= 0.05, 95% CrI = [0.02, 0.08], P($\hat{\beta}$ > 0) = 0.99), suggesting that the effect of Distractor Plausibility was much stronger in experiment 7 compared to the same effect in experiment 11. The interaction between Distractor Plausibility and Experiment had a credible interval of positive values only ($\hat{\beta}$= 0.14, 95% CrI = [0.10, 0.18], P($\hat{\beta}$ > 0) = 1).

The output of the model could explain the very strong effect of distractor plausibility observed in the critical region in experiment 7 (Maze Task II) and the same effect in the critical region in experiment 11 (Maze Task IV) a bit better. The comparison model that was run here shows that despite the effect being observed in both experiments, the individual differences (namely the different structure of the experimental sentences used in each experiment) either result in a much stronger effect (for experiment 7) or for not as strong distractor plausibility effects (in experiment 11). The model indicates that an increased structural difficulty in the sentences used in an

experiment leads to shorter reaction times suggesting a facilitation due to the plausibility of the distractor, while the effects are smaller and without such an increased facilitation (therefore longer reaction times) for the simpler structures. Overall then, the model run here shows that the different experiment and more specifically the associated differences in the structure of the sentences used in each of those experiments (as the task in both experiments was the same — maze task), determines the size of the effect of distractor plausibility.

3.11 General Discussion

The previous five main experiments (Experiment 7: Maze Task II, Experiment 8: Maze Task III, Experiment 9: Rapid Serial Visual Presentation I, Experiment 11: Maze Task IV, and Experiment 12: Rapid Serial Visual Presentation II) provided evidence that in the filler-gap sentences tested there were indeed semantic attraction effects in some cases. However to which extent those could be explained through cue-based retrieval models is not fully clear. At the same time, there were also some interesting findings which were rather unexpected.

3.11.1 Offline Effects

The offline results from the previous set of experiments did not show any particular semantic attraction effect.

No semantic attraction effect was observed in the accuracy in comprehension question responses in experiment 7 (maze task II). In line with the lack of an offline effect in the previous Cunnings and Sturt (2018) study, which was the basis of experiment 7, no effect of the distractor plausibility was seen. Despite the cue-based model not necessarily predicting attraction effects in offline processing, considering that those theories are focused on a so called race for activation and therefore more easily applied online (Cunnings & Sturt, 2018; Lewis & Vasishth, 2005; Nicenboim & Vasishth, 2018; Smith & Vasishth, 2020; Van Gompel et al., 2000), some studies have indeed found attraction effects in offline data and more so for ungrammatical sentences (i.e., Wagers et al., 2009). As such, an attraction effect could have been expected here, however nothing was found.

Some effects were also observed in experiment 9 (rapid serial visual presentation I) with attraction leading to sentences overall being judged as more acceptable when they included a distractor word that was a plausible complement for the verb in them.

3.11.2 Online Effects

The online results were partially compatible both with the expectations and hypotheses set at the start of this set of experiments, as well as with those by the cue-based retrieval accounts. The online measures in the previous five experiments showed that there were indeed semantic attraction effects in some of them.

In the three experiments involving a maze task (experiments 7, 8, and 11), semantic attraction effects were observed in the critical region (the verb) that interfered with the filler-gap dependencies' processing. In experiment 7 the online results were in agreement with the Cunnings and Sturt (2018). Strong semantic attraction effects

were observed in sentences independently of any local ambiguities that might have affected their processing.

In experiment 8 (maze task III) the results were of much greater interest. Semantic attraction effects were observed in that task too, leading to overall facilitation in sentences containing a plausible distractor word, echoing the results from the previous experiment as well as Cunnings and Sturt (2018). It is very notable that in this experiment, where participants had to make choices in sentences that they could built to be either plausible or implausible based on their choice for the last word (an example experimental item is given in table 3.7), attraction effects were manifested for overall plausible sentences too, with a considerable portion of participants selecting a continuation verb matching an intervening distractor word and not the correct grammatical word. Interestingly, the existence of a potentially plausible distractor word led to participants in some cases choosing a verb that was semantically compatible with those distractor words but would make the overall sentence implausible, while even when they chose the verb that was the correct continuation, the existence of a plausible distractor word in the sentence led to significantly slower responses, inhibiting processing.

Lastly, semantic attraction effects supporting the results of the previous studies were also seen in experiment 11 (maze task IV). The attraction findings in this experiment together with the comparison model run for the observations both in this and in experiment 7 indicate that semantic attraction effects in this set of experiments were modulated by the complexity of the sentence. That modulation of that effect is manifested to such a degree, that when in experiment 11 (the maze task IV that tested structurally simpler sentences) the attraction effects are observed in the critical region only leading to overall facilitated processing, in experiment 7 (which was the maze task II testing structurally more complex sentences) attraction effects were much stronger both in terms of the size of the effect as seen in the comparison model, as well as in the regions that the attraction effects were detected which were both in the critical and the spillover region. That is in line with what was originally hypothesised, as in that semantic attraction effects as the ones reported in Cunnings and Sturt (2018) could be attributed to the increased complexity of the tested sentences. Subjects' reaction times when tested in simpler filler-gap dependencies compared to when tested on more structurally demanding ones appear to confirm that hypothesis.

3.11.3 Theoretical Accounts

Cue-based Model

The findings from this set of experiments could partially be explained via a cue-based retrieval approach (Cunnings & Sturt, 2018; Lewis & Vasishth, 2005; Nicenboim &

Vasishth, 2018; Smith & Vasishth, 2020; Van Gompel et al., 2000). In those models, semantic attraction is an effect due to the match between the features that are stored in memory of an item that one encounters in a sentence, with the features at another place in the sentence (i.e., for example at the gap). Indeed the previous set of experiments showed that having a plausible distractor word in the sentences in question could lead to interference effects manifested as faster reaction times in experiment 7 and experiment 8, higher acceptance rates in experiment 9 and faster reaction times in experiment 11. Those results could be attributed to the existence of a word that semantically matched the verb in those sentences and that would be in line with the cue-based retrieval model, as the findings could be explained as support for interference effects based on semantic attraction.

What cannot be explained by a cue-based account is why semantic attraction effects were also present in sentences that could have been otherwise plausible in experiment 8 (maze task III) and why those attraction effects are modulated by the structure of the sentence. Such findings are not necessarily predicted by a cue-based model, as a cue-based model could not explain why participants would actively choose a continuation word that would make a sentence implausible even though they were given another plausible option. Especially if based on a cue-based retrieval model attraction is in itself a rescue strategy for otherwise implausible or ungrammatical sentences (**i.e.**; Lago et al., 2015; Laurinavichyute & von der Malsburg, 2022; Wagers et al., 2009). Potentially, that could be due to the task itself. Attempting to apply the cue-based model on the results in the three maze tasks (experiments 7, 8, & 11), a thought could be that since the maze forces for incremental parsing, it forces participants to over-rely on certain cues to quickly complete the task on hand. That way the results in those experiments could maybe be explained in a cue-based framework as in that the nature of the task required for the more recent cues (in terms of relative distance to the verb) of the distractor to be semi-activated (especially semantic/plausibility cues) and a quick match between those and the verb in the end to be achieved, leading to attraction effects across plausible and implausible sentences. However, there was nothing in the findings that clearly support something like that. At the same time, the cue-based model does not connect attraction effects to structure, so the findings of the effect being more robust in some of the experiments versus others could not be adequately explained by a cue-based retrieval account.

Final Conclusion

This second set of experiments showed that there are indeed some attraction effects that could very well explained and predicted by cue-based retrieval accounts (i.e., online effects of experiment 7), but there are also others that the cue-based model could

not account for (i.e., higher rate of incorrect continuation word choices in experiment 8). The findings overall showed that in the presence of a potentially plausible complement for the verb (i.e., a plausible intervening distractor word), the processing of that verb could be facilitated. That was manifested in shorter reaction times in some of the previous experiments. Importantly, that facilitation in processing was observed not to be an exclusive parsing method in otherwise implausible sentences, but in plausible sentences too (experiment 8). Attraction effects then seem to affect the processing of plausible sentences too, suggesting that semantic attraction might not be just a repair mechanism. At the same time, the previous results clearly showed that the presence of semantic attraction effects and even more importantly the size of an observed effect is highly modulated by the very structure of a filler-gap dependency sentence.

Overall, the findings from the previous experiments suggest that semantic attraction in filler-gap dependencies is a phenomenon that can be present and which can interfere with the resolution of such dependencies in reading. How this type of attraction effect could be predicted and further explained however, is not necessarily only through cue-based retrieval accounts but could be attributed to different approaches that will be further discussed in the next chapter (chapter 4).

Chapter 4

The central theme of this thesis has been the investigation of semantic similarity based retrieval interference phenomena and their interpretation. The two different chapters (chapter 2 and chapter 3) looked into those phenomena from a different prism. The set of experiments in chapter 2 looked at similarity based retrieval interference and semantic attraction from sentence external distractor words having as a starting point the Van Dyke and McElree (2006) study. The set of experiments in chapter 3 looked into similarity based retrieval interference effects and semantic attraction from sentence internal distractor words having as a starting point the interesting findings in the Cunnings and Sturt (2018) study.

4.1 Results and Cue-based Retrieval Models

The original studies on which both chapters 2 and 3 were based on, put high stock into attempting to explain their findings through cue based retrieval mechanisms, adopted and more computationally expressed by (Lewis & Vasishth, 2005; Lewis et al., 2006; McElree, 2000, 2006; McElree et al., 2003; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2011).

The findings in both chapters here though indicate that the cue based parsing mechanisms might not be the best parsing methods as of yet, at least for the type of dependencies studied here.

The issues that arise based on the findings in this thesis are two. One, that cue-based retrieval models do not take into high consideration the syntax and structural information of sentences. And two, that there needs to be a more clear idea as to whether cue-based retrieval that leads to semantic attraction effects is a result of rescuing otherwise problematic sentences.

In the investigation of semantic attraction effects from sentence external distractor words in chapter 2, there seems to be impossible to explain the findings through cue-based retrieval mechanisms (Lewis & Vasishth, 2005; Lewis et al., 2006; McElree, 2000, 2006; McElree et al., 2003; Van Dyke & Lewis, 2003; Van Dyke & McElree,

2011). Indeed the experiments in the chapter showed that in the resolution of filler-gap dependencies syntactic structure should be (at least partially) considered (**i.e.**; Clifton Jr & Frazier, 1989; Crain & Fodor, 1985; Ferreira & Clifton Jr, 1986; Fodor, 1978; Frazier, 1987a, 1987b; Frazier & d'Arcais, 1989; Frazier & Fodor, 1978; Gibson, 1998; Stabler, 2015; Stowe, 1986; Tanenhaus & Donnenwerth-Nolan, 1984). The proposed thematic integration account was heavily based on the notion that elements in language that share similar semantic cues lead to facilitation in processing (Badecker & Straub, 2002; Jäger et al., 2015; Pearlmutter et al., 1999; Van Dyke & McElree, 2006). That is where the element of relatedness came from in the account. Elements that are closely related and share many semantic cues are more likely to lead to faster / easier parsing. However, the thematic integration account was equally based on the syntactic structure of the sentence and the specific structural characteristics of the filler-gap dependencies. Interference in processing should only have been expected if all elements that are considered as potential arguments of a verb in a sentence are at least in a structure and not just words that have no structural role.

Overall then, from chapter 2 it seems that syntax should be accommodated in some way in those models by incorporating structure as an equal driving force during parsing. Together with the lexical association, the semantic or any other feature-cues that cue-based retrieval models could include, having features that are bound by the structural position of where an item is (at the very least in the same sentence in the experiments described in chapter 2), seem to play a role in parsing. Also by implementing at the very least that a cue-based retrieval interference effect (more so and one that interferes with the integration of long distance dependencies (**i.e**; Aoshima et al., 2004; Garnsey et al., 1989; Grodner & Gibson, 2005; Omaki et al., 2014; Phillips et al., 2005; M. Pickering & Barry, 1991; Stowe, 1986)) is caused by elements that have some structural features stored in their chunk of cues.

The investigation of semantic attraction phenomena from sentence internal distractors indeed showed that the cue-based retrieval models could make some accurate predictions. The sharing of semantic cues between the retrieval gap in a filler gap dependency and an intervening distractor element does indeed lead to facilitated processing. That is primarily for implausible sentences in line with other studies that have observed such effects in similarly overall implausible sentences (Kuperberg, 2007; Levy, 2008; Van Petten & Luka, 2012). The cue-based retrieval models postulate that such an effect is in place as the parser attempt to *save* the otherwise problematic dependency. Interestingly, a finding that could not necessarily be predicted in the current form of the cue-based retrieval models then, is that semantic attraction interference was also found for plausible sentences (Lago et al., 2015). The findings of experiment 8 (maze task III) questioned why participants would actively choose an ungrammatical option that would make for implausible sentences if the cue-based

models consider those effect to occur in implausible sentences.

On top of that, the need for an implementation in the cue-based model of syntactic complexity was also evident in the experiments in chapter 3 too. For some of the experiments, the effects might have been easily predicted by a cue-based model, however what is not predicted via such a mechanism is the importance of syntax when it comes to the size of that interference effect. The experiments alone as well as the model comparison between them showed that the presence of an interference effect and how big or small that is, is modulated by the overall complexity of the sentence.

Chapter 3 then mainly suggests that there should be an implementation in the cue-based models as to what kind of information is truly involved in parsing and if attraction is indeed a rescue strategy. And it once again in agreement with the findings in chapter 2 underlines the need for including syntactic structural information in the cue-based model.

From the previous findings in chapters 2 and 3, it seems that while interference is a phenomenon that can better explain some effects in the psycholinguistic literature (Cunnings & Sturt, 2014, 2018; Fedorenko et al., 2006; Gordon et al., 2002; Van Dyke, 2007; Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006, 2011; Wagers et al., 2009), there needs to be some implementations potentially inspired by other accounts that also prioritise structural information (i.e., active filler strategy, minimal chain principle, dependency locality theory, etc.).

Lastly, cue-based models can indeed give great predictions when it comes to linguistics phenomena. Importantly though, their predictions should be more tailored to language. As syntax and structure are at the very basis of language, I believe that those should also be implemented together with other concepts that at the very forefront of those models (like semantic similarity, frequency, etc.) so that they can better fit linguistic phenomena.

4.2 Effect of Task

Following from the previous discussion on whether cue-based models predict mistakes in plausibility, another interesting finding to consider is whether that interesting effect was task-induced.

Considering that the maze task is a task that forces incremental parsing (Boyce et al., 2020; Forster et al., 2009; Witzel et al., 2012), as one makes choices from one word to the next, it is intriguing to think whether the task actually induced those effects with participants choosing an overall implausible continuation in the presence of a plausible distractor word.

One might think that the maze task itself taps into processes that combine both

online processing and comprehension but also production (Forster et al., 2009; Witzel et al., 2012). Despite not having any other such effects in the rest of the experiments run with the maze task methodology, participants tapping into production processes could potentially explain that finding.

However in other experiments that tested items previously run with other methodologies (i.e., eye-tracking in Cunnings and Sturt (2018)), the same effects were replicated with a high similarity between the observed effects in the maze task and the same effects in later measures of fixations (i.e., total fixation times). Consequently, the question of whether production is involved at all, is not clear at this stage and based on the particular experiments run.

4.3 A Different Perspective

Away from the cue-based retrieval models, a potential account for the found effects in the presented experiments is offered by some predictive processing models (Chow et al., 2016; Liao et al., 2022), where the distractor noun generates certain predictions for the main verb, which together with its closer proximity has a greater impact on the experiments presented.

Chow et al. (2016) suggested the "Bag of arguments" hypothesis, in it meaning that structure already impacts prediction even before a subset of noun phrases are identified as arguments of an upcoming verb, while importantly the prediction of that verb is also constrained by that information. Similar to that another hypothesis was suggested, the "Bag of words" one. The main difference between the two being that the bag of words relies on quick feature association effects, while the bag of arguments also assumes structural constraints in those early associative effects.

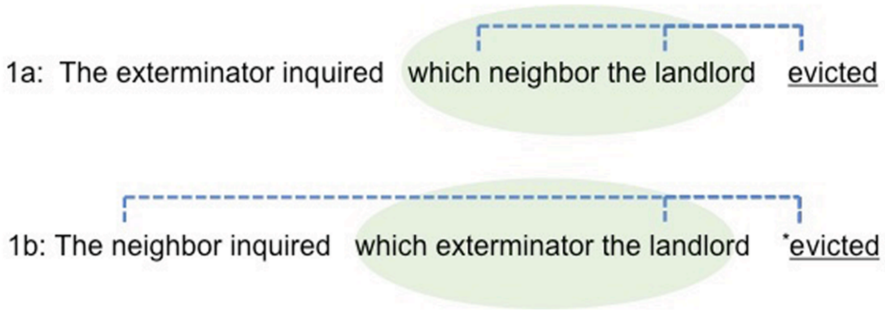


Figure 4.1: *Visual illustrations for stimuli in Chow et al. (2016). Dotted line indicates semantic associations and color shading shows argument positions of the embedded verb. The illustration is taken by Liao et al. (2022)*

Testing sentences like the ones in figure 4.1 in ERP, Chow et al. (2016) found larger N400 response at the embedded verb in sentences like those in 1b compared to those

in 1a. They took that to confirm the bag of arguments hypothesis, as facilitative effects on the verb from semantic associates were larger when these associates were in argument positions of the verb (sentence in 1a), compared to a case where one of the associates appears in a non-argument position (as in 1b).

Such a perspective could explain the effects in most of the experiments run for this thesis. In the first set of experiments (chapter 2), such an account could explain why the observed facilitation effects were due to the semantic relationship between a word that was in an argument position in the sentence and ultimately in closer proximity, whereas words that were sentence external distractors and therefore had no argument position and were further, could not affect processing. In the second set of experiments (chapter 3), the much closer proximity of the distractor word together with the close semantic association between plausible distractor words and verb, could explain not only the interference effects in implausible sentences, but also the interesting finding in experiment 8 (maze task III), where participants chose a verb that would make a sentence globally implausible in higher rates when preceded by a plausible distractor word.

4.4 Future Directions & Conclusion

Based on the previous experiments and their findings, future work could involve a closer investigation of the phenomenon in experiment 8. A clever design that would combine more sensitive measures, such as eye-tracking or ERPs together with an open option where participants can actively choose between plausible and implausible continuations of dependencies would ideally allow for a clearer idea on why that effect was observed and its actual size.

Furthermore, it would be interesting to test similar sentences with varying levels of plausibility in an ERP design, in order to test whether the bag of arguments or even the bag of words can explain those phenomena better than the accounts already discussed (i.e., cue-based models).

More work also needs to be done on implementing the otherwise very useful and accurate cue-based retrieval model, so that they can make better predictions on several linguistic phenomena.

In conclusion, the resolution of filler-gap dependencies appears to be occasionally affected by intervening distractor words on the basis of semantic features. That is not the case though for sentence external distractor words, as there are some structural constraints in place that cannot be overlooked. Cue-based retrieval models can account for some semantic attraction effects in filler-gap dependencies, they need however to include structural feature information as well.

Chapter 5

Appendices

5.1 Appendix for tables

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Relatedness	0.76	[0.50, 1.02]

Table 5.1: *The estimated mean and the 95% highest posterior density interval for the fixed effect in the surprisal model for experiment 1 (Cloze Task).*

Fixed Effects	Estimate ($\hat{\beta}$)	*HPDI*
Related	-55.18	[-100.72, -12.26]
Unrelated	-30.71	[-75.05, 13.13]

Table 5.2: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the reaction time model for experiment 2 (Lexical Decision Task).*

	Critical region		**Spillover region**	
Fixed Effects	Estimate ($\hat{\beta}$)	*HPDI*	Estimate ($\hat{\beta}$)	*HPDI*
Load Type	-0.01	[-0.02, 0.01]	0.00	[-0.02, 0.02]
Verb Type	0.00	[-0.02, 0.02]	-0.00	[-0.02, 0.02]
Load Type$\times$*Verb Type*	0.01	[-0.01, 0.02]	0.01	[-0.01, 0.03]

Table 5.3: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the models for the reading times at the critical and at the spillover regions in experiment 3 (Dual Task Self-paced Reading). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effects	Estimate ($\hat{\beta}$)	*HPDI*
Load Type	-0.13	[-0.31, 0.05]
Verb Type	-0.50	[-0.93, -0.09]
Load Type×*Verb Type*	0.04	[-0.14, 0.22]

Table 5.4: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the question accuracy model for experiment 3 (Dual Task Self-paced Reading). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effects	Estimate ($\hat{\beta}$)	*HPDI*
Load Type	0.01	[-0.00, 0.02]
Verb Type	0.04	[0.02, 0.06]
Load Type×*Verb Type*	0.00	[-0.01, 0.01]

Table 5.5: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the question response times model in experiment 3 (Dual Task Self-paced Reading). The estimated mean and the 95% credible intervals are on log scale.*

	Critical region		**Spillover region**	
Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*	Estimate ($\hat{\beta}$)	*HPDI*
Relatedness	0.09	[0.03, 0.15]	0.03	[-0.01, 0.07]

Table 5.6: *The estimated mean and the 95% highest posterior density interval for the fixed effect in the models for the reading times at the critical and at the spillover regions in experiment 4 (Maze Task I). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Relatedness	-0.86	[-1.31, -0.42]

Table 5.7: *The estimated mean and the 95% highest posterior density interval for the fixed effect in the question accuracy model in experiment 4 (Maze Task I). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Relatedness	0.08	[0.01, 0.15]

Table 5.8: *The estimated mean and the 95% highest posterior density interval for the fixed effect in the question response times model in experiment 4 (Maze Task I). The estimated mean and the 95% credible intervals are on log scale.*

	Critical region		**Spillover region**	
Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*	Estimate ($\hat{\beta}$)	*HPDI*
Relatedness	0.00	[-0.03, 0.03]	0.03	[0.00, 0.07]

Table 5.9: *The estimated mean and the 95% highest posterior density interval for the fixed effect in the model for the reading times at the critical and at the spillover regions in experiment 5 (Simple Self-paced Reading Task). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Relatedness	-0.74	[-1.43, -0.05]

Table 5.10: *The estimated mean and the 95% highest posterior density interval for the fixed effect in the question accuracy model in experiment 5 (Simple Self-paced Reading Task). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Relatedness	0.12	[0.06, 0.18]

Table 5.11: *The estimated mean and the 95% highest posterior density interval for the fixed effect in the question response times model in experiment 5 (Simple Self-paced Reading Task). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Experiment	-1.09	[-1.13, -1.05]
Relatedness	0.09	[0.04, 0.14]
Experiment×*Relatedness*	-0.09	[-0.14, -0.04]

Table 5.12: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the comparison model for experiments 4 and 5. The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Verb Ambiguity	-0.68	[-1.67, 0.18]
Distractor Plausibility	-0.58	[-1.50, 0.27]
Verb Ambiguity×*Distractor Plausibility*	0.87	[0.03, 1.83]

Table 5.13: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the accuracy in selection model in experiment 7 (Maze Task II). The estimated mean and the 95% credible intervals are on log scale.*

	Critical region		**Spillover region**	
Fixed Effects	Estimate ($\hat{\beta}$)	*HPDI*	Estimate ($\hat{\beta}$)	*HPDI*
Verb Ambiguity	-0.01	[-0.03, 0.01]	0.00	[-0.02, 0.02]
Distractor Plausibility	0.09	[0.07, 0.12]	0.04	[0.01, 0.06]
Verb Ambiguity×*Distractor Plausibility*	-0.00	[-0.02, 0.02]	-0.00	[-0.02, 0.02]

Table 5.14: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the models for the response times at the critical and at the spillover regions in experiment 7 (Maze Task II). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Distractor Plausibility	0.71	[0.57, 0.86]

Table 5.15: *The estimated mean and the 95% highest posterior density interval for the fixed effect in the accuracy in selection model in experiment 8 (Maze Task III). The estimated mean and the 95% credible intervals are on log scale.*

	Critical region	
Fixed Effects	Estimate ($\hat{\beta}$)	*HPDI*
Verb Choice	0.02	[-0.01, 0.06]
Distractor Plausibility	0.05	[0.03, 0.07]
Verb Choice×*Distractor Plausibility*	0.07	[0.05, 0.09]

Table 5.16: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the model for the response times at the critical region in experiment 8 (Maze Task III). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Sentence Plausibility	0.42	[0.28, 0.56]
Distractor Plausibility	-0.23	[-0.49, 0.01]
Sentence Plausibility×*Distractor Plausibility*	-0.25	[-0.38, -0.13]

Table 5.17: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the accuracy model in experiment 9 (Rapid Serial Visual Presentation Task I). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Sentence Plausibility	-0.04	[-0.07, -0.02]
Distractor Plausibility	-0.02	[-0.06, 0.02]
Sentence Plausibility×*Distractor Plausibility*	0.02	[-0.02, 0.05]

Table 5.18: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the response times model in experiment 9 (Rapid Serial Visual Presentation Task I). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Sentence Plausibility	-1.56	[-2.91, -0.14]
Distractor Plausibility	9.09	[7.13, 11.07]
Sentence Plausibility×*Distractor Plausibility*	3.23	[1.72, 4.59]

Table 5.19: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the acceptability rates model in experiment 9 (Rapid Serial Visual Presentation Task I). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Noun Plausibility	-0.26	[-0.95, 0.42]
Distractor Plausibility	0.69	[0.02, 1.43]
Noun Plausibility×*Distractor Plausibility*	0.26	[-0.39, 0.97]

Table 5.20: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the accuracy in selection model in experiment 11 (Maze Task IV). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effects	Critical region		Spillover region	
	Estimate ($\hat{\beta}$)	*HPDI*	Estimate ($\hat{\beta}$)	*HPDI*
Noun Plausibility	-0.09	[-0.12, -0.06]	0.04	[0.03, 0.06]
Distractor Plausibility	0.03	[0.01, 0.04]	-0.01	[-0.02, 0.01]
Noun Plausibility×*Distractor Plausibility*	-0.00	[-0.01, 0.01]	-0.00	[-0.01, 0.02]

Table 5.21: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the models for the response times at the critical and at the spillover regions in experiment 11 (Maze Task IV). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Subject Noun	0.63	[0.29, 0.96]
Distractor Plausibility	0.04	[-0.09, 0.18]
Subject Noun×*Distractor Plausibility*	-0.26	[-0.45, -0.08]

Table 5.22: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the accuracy model in experiment 12 (Rapid Serial Visual Presentation Task II). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Subject Noun	-0.06	[-0.10, -0.02]
Distractor Plausibility	0.00	[-0.02, 0.03]
Subject Noun×*Distractor Plausibility*	0.03	[0.01, 0.06]

Table 5.23: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the response times model in experiment 12 (Rapid Serial Visual Presentation Task II). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Subject Noun	9.35	[7.43, 11.22]
Distractor Plausibility	-1.39	[-2.77, 0.04]
Subject Noun×*Distractor Plausibility*	1.10	[-0.10, 2.32]

Table 5.24: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the acceptability rates model in experiment 12 (Rapid Serial Visual Presentation Task II). The estimated mean and the 95% credible intervals are on log scale.*

Fixed Effect	Estimate ($\hat{\beta}$)	*HPDI*
Experiment	-0.13	[-0.16, -0.09]
Distractor Plausibility	0.05	[0.02, 0.08]
Experiment×*Distractor Plausibility*	0.14	[0.10, 0.18]

Table 5.25: *The estimated mean and the 95% highest posterior density interval for the fixed effects in the comparison model for experiments 7 and 11. The estimated mean and the 95% credible intervals are on log scale.*

5.2 Appendix for graphs



Figure 5.1: Posterior distributions of the fixed effects parameter in the surprisal model in experiment 1. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

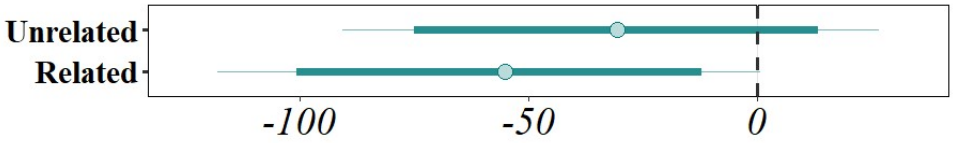


Figure 5.2: Posterior distributions of the fixed effects parameters in the priming model in experiment 2. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.



Figure 5.3: Posterior distributions of the fixed effects parameters in the comprehension question accuracy model in experiment 3. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

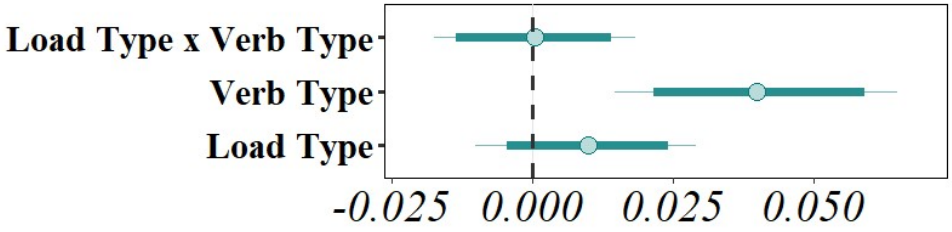


Figure 5.4: Posterior distributions of the fixed effects parameters in the comprehension question response times model in experiment 3. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

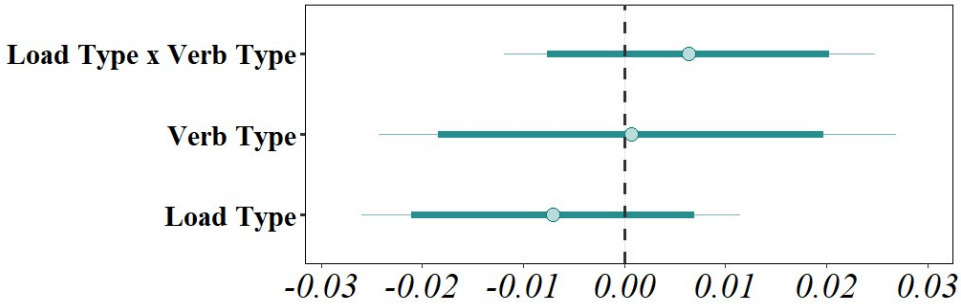


Figure 5.5: Posterior distributions of the fixed effects parameters in the reading times at the critical region model in experiment 3. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

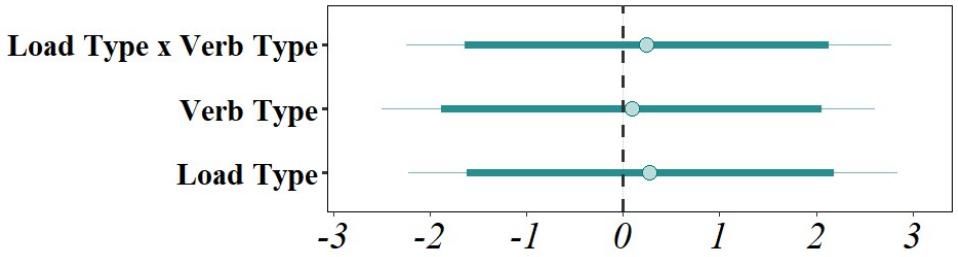


Figure 5.6: Posterior distributions of the fixed effects parameters in the reading times at the spillover region in experiment 3. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

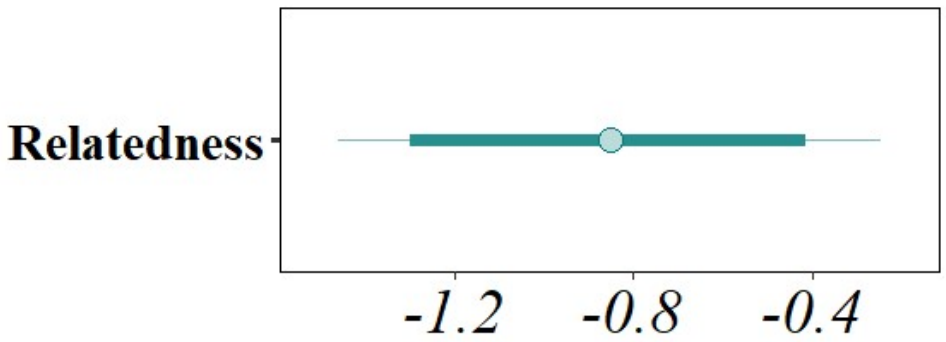


Figure 5.7: Posterior distributions of the fixed effects parameter in the comprehension question accuracy model in experiment 4. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

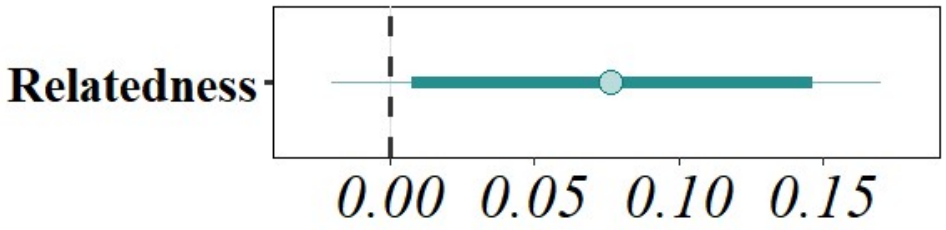


Figure 5.8: Posterior distributions of the fixed effects parameter in the comprehension question response times model in experiment 4. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

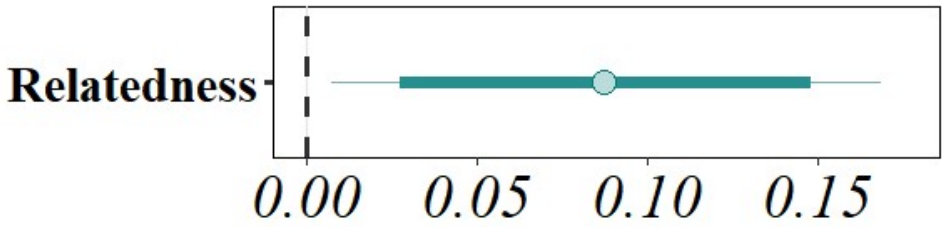


Figure 5.9: Posterior distributions of the fixed effects parameter in the reading times at the critical region model in experiment 4. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

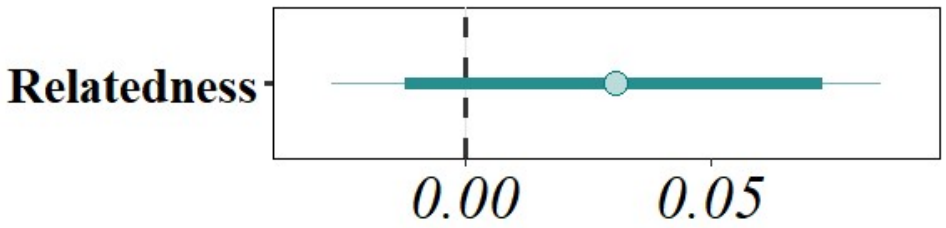


Figure 5.10: Posterior distributions of the fixed effects parameter in the reading times at the spillover region model in experiment 4. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

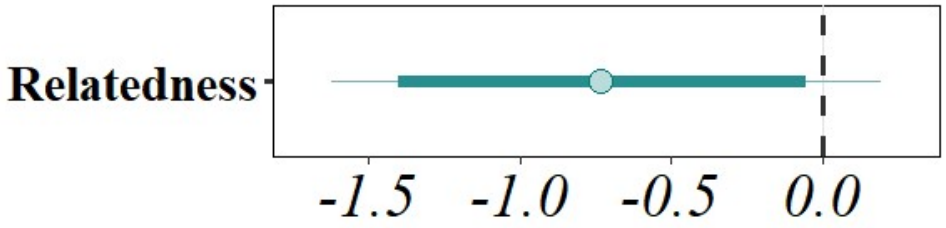


Figure 5.11: Posterior distributions of the fixed effects parameter in the comprehension question accuracy model in experiment 5. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

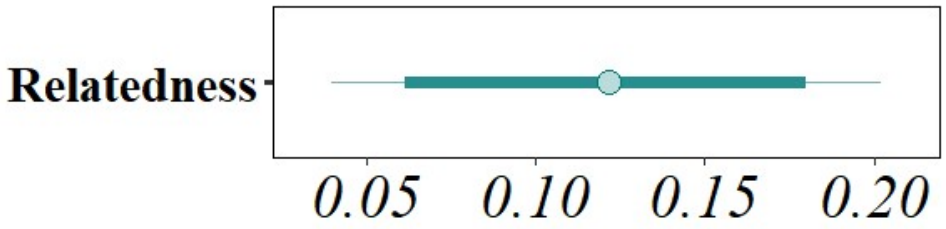


Figure 5.12: Posterior distributions of the fixed effects parameter in the comprehension question response times model in experiment 5. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

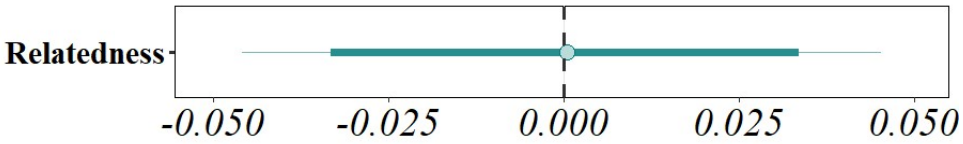


Figure 5.13: Posterior distributions of the fixed effects parameter in the reading times at the critical region model in experiment 5. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.



Figure 5.14: Posterior distributions of the fixed effects parameter in the reading times at the spillover region model in experiment 5. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

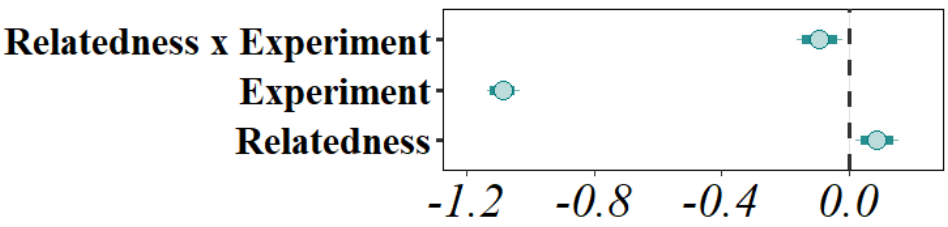


Figure 5.15: Posterior distributions of the fixed effects parameters in the comparison model for experiments 4 and 5. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

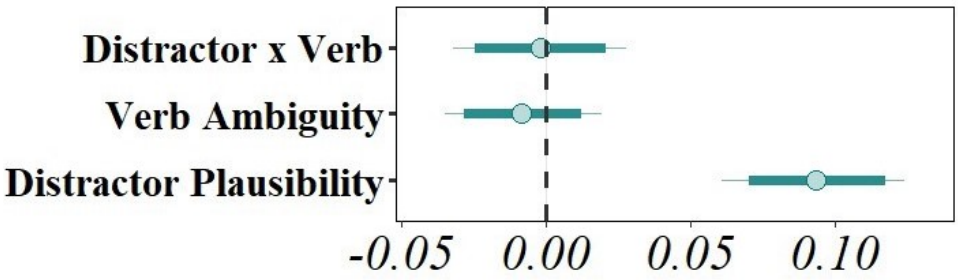


Figure 5.16: Posterior distributions of the fixed effects parameters in the response times at the critical region model in experiment 7. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

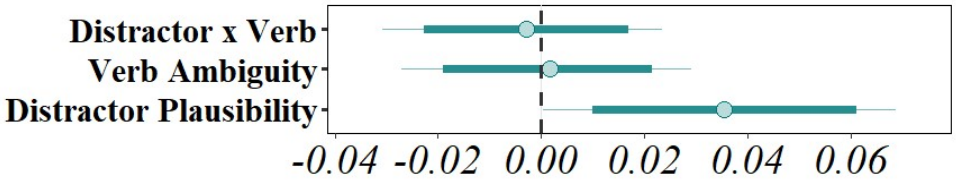


Figure 5.17: Posterior distributions of the fixed effects parameters in the response times at the spillover region model in experiment 7. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

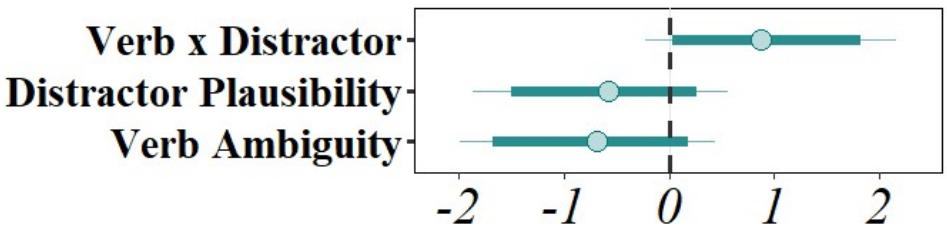


Figure 5.18: Posterior distributions of the fixed effects parameters in the accuracy at the critical region model in experiment 7. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

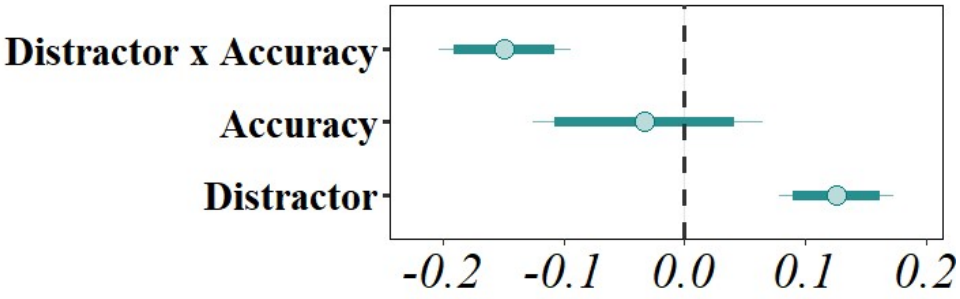


Figure 5.19: Posterior distributions of the fixed effects parameter in the response times at the critical region model in experiment 8. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

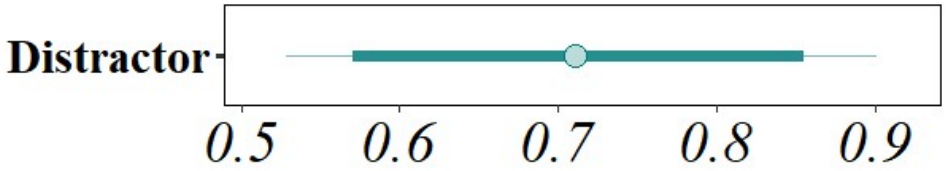


Figure 5.20: Posterior distributions of the fixed effects parameter in the accuracy at the critical region model in experiment 8. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

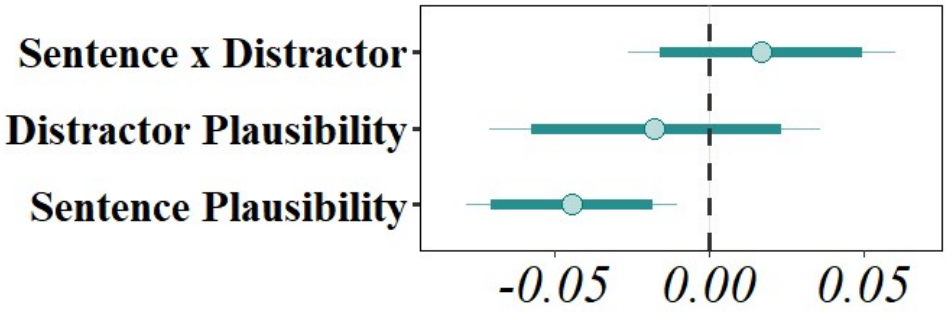


Figure 5.21: Posterior distributions of the fixed effects parameters in the response times model in experiment 9. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

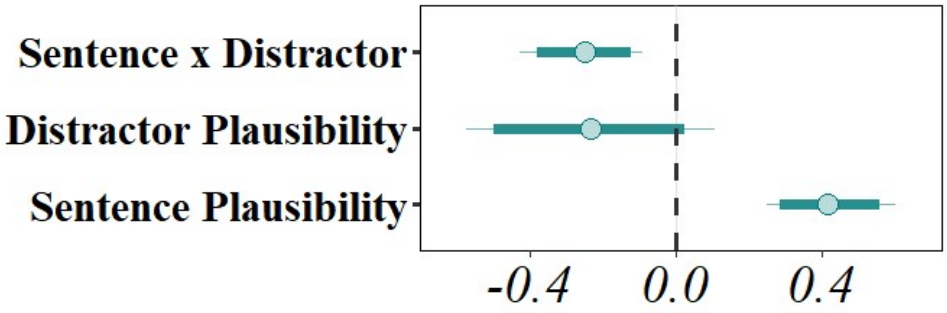


Figure 5.22: Posterior distributions of the fixed effects parameters in the accuracy model in experiment 9. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

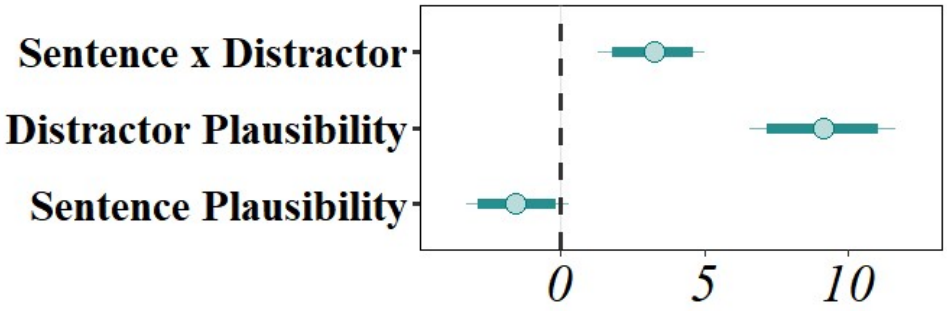


Figure 5.23: Posterior distributions of the fixed effects parameters in the acceptability rates model in experiment 9. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

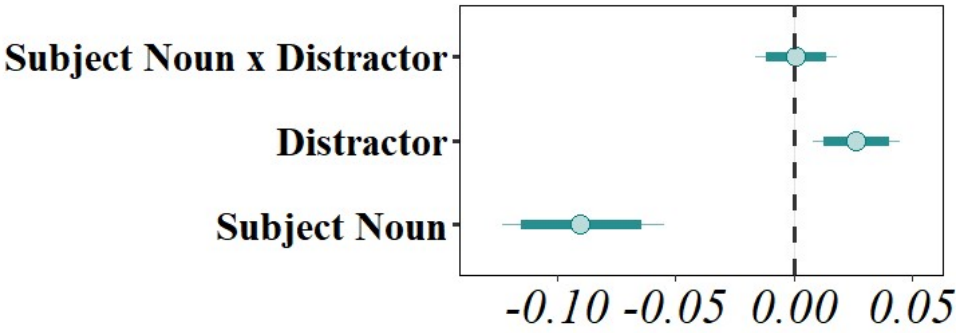


Figure 5.24: Posterior distributions of the fixed effects parameters in the response times at the critical region model in experiment 11. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

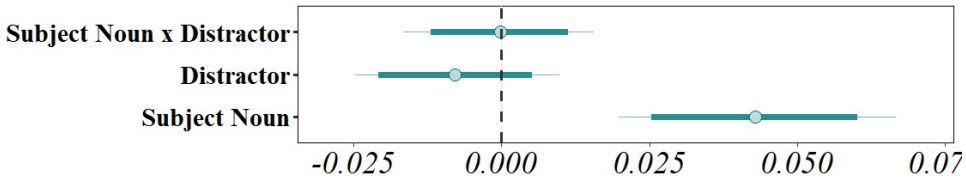


Figure 5.25: Posterior distributions of the fixed effects parameters in the response times at the spillover region model in experiment 11. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

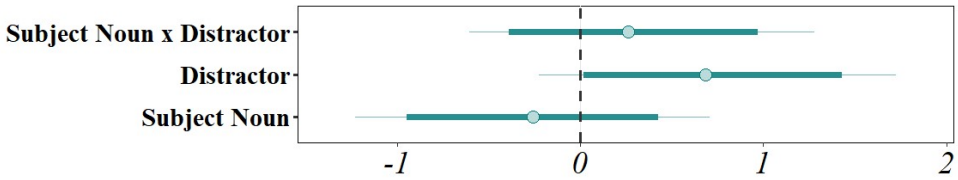


Figure 5.26: Posterior distributions of the fixed effects parameters in the accuracy at the critical region model in experiment 11. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

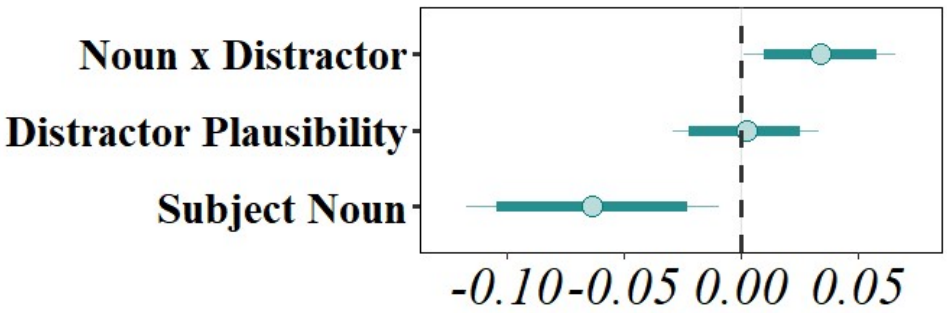


Figure 5.27: Posterior distributions of the fixed effects parameters in the response times model in experiment 12. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

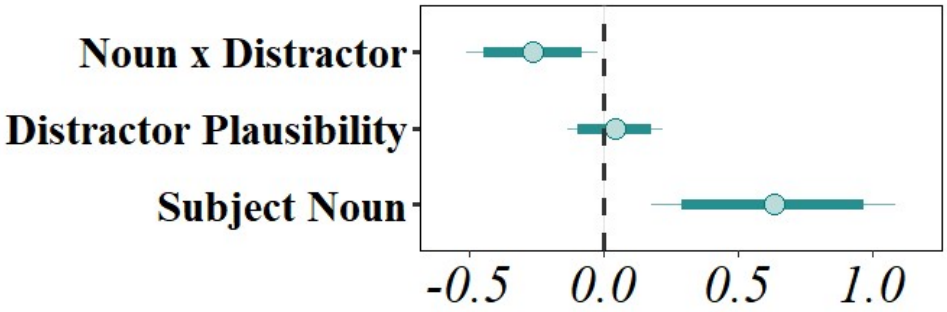


Figure 5.28: Posterior distributions of the fixed effects parameters in the accuracy model in experiment 12. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

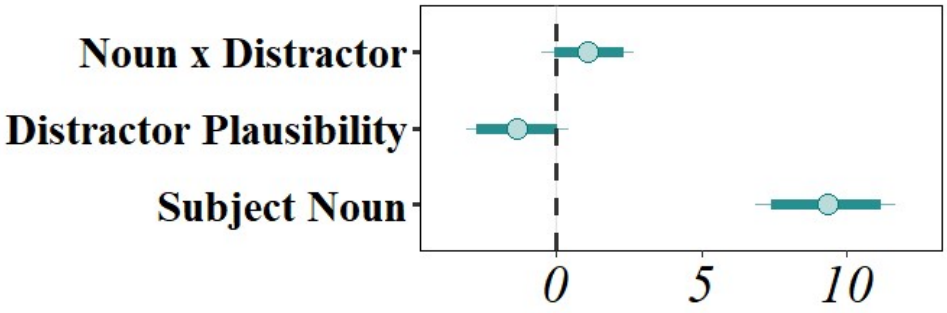


Figure 5.29: Posterior distributions of the fixed effects parameters in the acceptability rates model in experiment 12. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

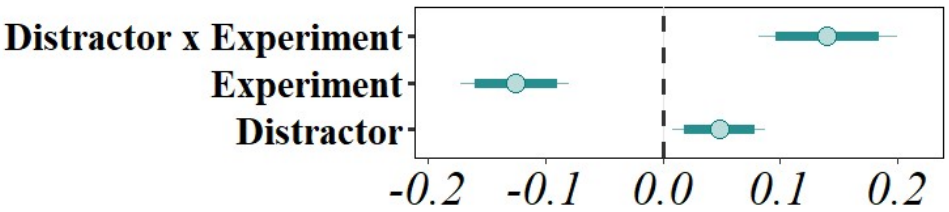


Figure 5.30: Posterior distributions of the fixed effects parameters in the comparison model for experiments 7 and 11. The dark blue line marks the posterior mean, the light blue shaded area marks the 95% credible intervals and zero is marked with a vertical dashed line.

5.3 Appendix for experimental stimuli

This appendix contains the experimental stimuli used in the main experiments contained in this thesis (experiments 3, 4, 5, 7, 8, 9, 11, & 12).

5.3.1 Experiment 3 stimuli

Below are the experimental items from Experiment 3. Each item consists of the memory load (nominal / verbal), the related condition in (a), the unrelated condition in (b) and the comprehension question followed by Y, if the correct answer was yes or N, if the correct answer was no.

1. table-sink-truck/chop-learn-try

   (a) It was the boat that the guy who lived by the sea sailed in two sunny days.

   (b) It was the boat that the guy who lived by the sea fixed in two sunny days.

   Did the guy live by the sea?,Y

2. mug-money-prize/draw-wiggle-burn

   (a) It was the dress that the waitress who worked at the diner wore on the last Monday of the month.

   (b) It was the dress that the waitress who worked at the diner won on the last Monday of the month.

   Did the waitress wear the dress on Wednesday?,N

3. gown-poster-room/stir-shoot-clean

   (a) It was the makeup that the performer who arrived very late applied before the stage appearance.

   (b) It was the makeup that the performer who arrived very late ruined before the stage appearance.

   Did the performer apply the makeup before the stage appearance?,Y

4. sack-art-silver/forget-float-pour

   (a) It was the gun that the character who began the fight shot after the Western sunset.

   (b) It was the gun that the character who began the fight seized after the Western sunset.

Was it the film that the character began?,N

5. beverage-wastebasket-clothing/cater-fret-collect

   (a) It was the cigarette that the criminal who robbed the electronics store smoked in the dark cell.

   (b) It was the cigarette that the criminal who robbed the electronics store sought in the dark cell.

   Was it the cell in which the criminal smoked the cigarette?,Y

6. key-pen-earring/injure-scatter-arise

   (a) It was the button that the maid who returned from vacation sewed in the early morning.

   (b) It was the button that the maid who returned from vacation spotted in the early morning.

   Was the button sewed in the evening?,N

7. musket-tent-drink/weigh-discuss-close

   (a) It was the gum that the camper who forgot to bring sandwiches chewed during the forest campfire.

   (b) It was the gum that the camper who forgot to bring sandwiches fetched during the forest campfire.

   Did the camper forget to bring sandwiches?,Y

8. package-apologies-bouquet/grieve-use-bake

   (a) It was the bill that the citizen who overslept for work paid after the phone call.

   (b) It was the bill that the citizen who overslept for work sent after the phone call.

   Did the citizen pay the bill before the phone call?,N

9. haircut-snow-costume/arrive-steal-expel

   (a) It was the solo that the singer who practiced for months performed throughout the entire evening.

   (b) It was the solo that the singer who practiced for months dreaded throughout the entire evening.

Did the singer perform the solo?,Y

10. doll-tie-note/hold-oppose-reach

    (a) It was the puzzle that the lad who received good grades solved after the mediocre lunch.

    (b) It was the puzzle that the lad who received good grades mislaid after the mediocre lunch.

    Was the lad's lunch delicious?,N

11. silverware-shovel-flute/lose-indicate-eat

    (a) It was the perfume that the passerby who shopped for presents smelled during the holiday season.

    (b) It was the perfume that the passerby who shopped for presents stole during the holiday season.

    Did the passerby shop for presents?,Y

12. laptop-photograph-video/build-dig-choose

    (a) It was the hair that the student who rented a limo shampooed before her prom night.

    (b) It was the hair that the student who rented a limo damaged before her prom night.

    Was the limo rented by the teacher?,N

13. shirt-cup-floor/know-care-teach

    (a) It was the car that the husband who changed his occupation drove after the delicious dinner .

    (b) It was the car that the husband who changed his occupation washed after the delicious dinner.

    Did the husband drive the car?,Y

14. clause-caption-assignment/digress-shout-finish

    (a) It was the number that the secretary who collected the commission dialed during her break.

    (b) It was the number that the secretary who collected the commission memorised during her break.

Did the secretary dial the number after her break?,N

15. soup-wreath-stockings/stroke-believe-seem

    (a) It was the rug that the house attendant who quit his previous job swept on the rainy Wednesday.

    (b) It was the rug that the house attendant who quit his previous job destroyed on the rainy Wednesday.

    Was it the rug that was swept?,Y

16. entrance-safe-cabinet/die-capture-give

    (a) It was the museum that the philanthropist who journeyed to South America founded in the middle of his vacation.

    (b) It was the museum that the philanthropist who journeyed to South America accessed in the middle of his vacation.

    Did the philanthropist journey to Africa?,N

17. coat-medicine-racket/crave-ask-irritate

    (a) It was the newspaper that the assistant who stayed after hours printed from the headquarters.

    (b) It was the newspaper that the assistant who stayed after hours removed from the headquarters.

    Did the assistant print the newspaper?,Y

18. comb-bag-mask/explore-sprint-tend

    (a) It was the food that the son who entered middle school cooked on the Fourth of July.

    (b) It was the food that the son who entered middle school dropped on the Fourth of July.

    Was it the daughter who entered middle school?,N

19. chart-serum-software/shake-beckon-yell

    (a) It was the syringe that the doctor who opened a practice sterilized before the critical operation.

    (b) It was the syringe that the doctor who opened a practice misused before the critical operation.

Was the syringe sterilized before the operation?,Y

20. van-office-studio/think-drive-swear

    (a) It was the suit that the accountant who acquired a date ironed at the local store.

    (b) It was the suit that the accountant who acquired a date rented at the local store.

    Did the accountant iron the suit at his sister's store?,N

21. iron-lock-crib/succeed-count-speak

    (a) It was the whistle that the boy who listened to his mother blew during the splendid celebration.

    (b) It was the whistle that the boy who listened to his mother damaged during the splendid celebration.

    Did the boy listen to his mother?,Y

22. oven-gadget-uniform/bleed-invest-have

    (a) It was the pie that the chef who moved to a new town baked for the cooking contest.

    (b) It was the pie that the chef who moved to a new town invented for the cooking contest.

    Was it the chef who baked the pie?,N

23. napkin-receipt-brochure/conjure-sleep-oblige

    (a) It was the card that the tourist who obtained the discount activated by the movie theatre.

    (b) It was the card that the tourist who obtained the discount discarded by the movie theatre.

    Was the card activated by the movie theatre?,Y

24. cupcake-dictionary-razor/admire-find-shirk

    (a) It was the jacket that the steward who crossed the world unzipped prior to the trip.

    (b) It was the jacket that the steward who crossed the world packed prior to the trip.

Was the world crossed by the pilot?,N

25. highlighter-snack-headphones/see-grow-move

    (a) It was the report that the writer who created the show typed after the afternoon broadcast.

    (b) It was the report that the writer who created the show misplaced after the afternoon broadcast.

    Was it the writer that typed the report?,Y

26. rum-garments-corn/bend-wink-sit

    (a) It was the hardware that the associate who mastered computer science programmed at the large organization.

    (b) It was the hardware that the associate who mastered computer science exported at the large organization.

    Did the associate program the software?,N

27. tea-furniture-scooter/play-digest-postpone

    (a) It was the book that the lady who attended church last Sunday read after the business meeting.

    (b) It was the book that the lady who attended church last Sunday sold after the business meeting.

    Did the lady attend church last Sunday?,Y

28. turtleneck-bowl-toothbrush/illustrate-touch-sing

    (a) It was the orange that the traveler who lost his luggage peeled after the delayed departure.

    (b) It was the orange that the traveler who lost his luggage rinsed after the delayed departure.

    Was it the pear that the traveler peeled?,N

29. sky-picture-fountain/open-compel-whisper

    (a) It was the flower that the wife who purchased the new house planted after the torrential rain.

    (b) It was the flower that the wife who purchased the new house admired after the torrential rain.

Was the flower planted after the rain?,Y

30. bathroom-headlight-trunk/cry-hear-freeze

    (a) It was the clutch that the driver who worked late nights pressed before the next shift.

    (b) It was the clutch that the driver who worked late nights repaired before the next shift.

    Did the driver press the brake?,N

31. vase-violin-flannel/like-expect-climb

    (a) It was the alcohol that the girl who won the sweepstakes poured before the rock concert.

    (b) It was the alcohol that the girl who won the sweepstakes purchased before the rock concert.

    Did the girl win the sweepstakes?,Y

32. game-limousine-mailbox/hate-keep-admit

    (a) It was the plane that the brother who moved to Arizona aviated until the unfortunate crash.

    (b) It was the plane that the brother who moved to Arizona owned until the unfortunate crash.

    Was it to Los Angeles that the brother moved?,N

33. plate-goggles-contacts/caress-become-pester

    (a) It was the code that the policeman who received a bonus entered in the shabby apartment.

    (b) It was the code that the policeman who received a bonus cracked in the shabby apartment.

    Was it the policeman who entered the code?,Y

34. disk-sport-cymbal/annul-slink-argue

    (a) It was the radio that the adolescent who finished his homework unplugged in the upstairs living room.

    (b) It was the radio that the adolescent who finished his homework played in the upstairs living room.

Was the radio unplugged in the bedroom?,N

35. purse-glasses-folder/do-write-consume

    (a) It was the drug that the physician who graduated last month injected before the next patient.

    (b) It was the drug that the physician who graduated last month misplaced before the next patient.

    Did the physician graduate last month?,Y

36. jeep-bike-computer/say-wonder-buy

    (a) It was the kite that the youth who turned twelve years old flew after his birthday party.

    (b) It was the kite that the youth who turned twelve years old crashed after his birthday party.

    Was the kite purchased by the youth?,N

5.3.2 Experiment 4 stimuli

Below are the experimental items from Experiment 4. Each item consists of the generated foil, the unrelated condition in (a), the related condition in (b) and the comprehension question followed by Y, if the correct answer was yes or N, if the correct answer was no.

1. x-x-x eye etc pray yeah pain hate drug youth map bet pre comics sir gay rabbi send.

   (a) It was the boat that the guy who lived by the sea fixed in two sunny days.

   (b) It was the boat that the guy who lived by the sea sailed in two sunny days.

   Did the guy live by the sea?,Y

2. x-x-x dad mom ought glad miss deleting mid recent al etc equip hate else hour yeah Insane uses risk goals.

   (a) It was the dress that the waitress who worked at the diner won on the last Monday of the month.

   (b) It was the dress that the waitress who worked at the diner wore on the last Monday of the month.

   Did the waitress wear the dress on Wednesday?, N

3. x-x-x hell kid closes feet etc discounts mid eastern hate rate wisdom forget feel yours republican.

   (a) It was the makeup that the performer who arrived very late ruined before the stage appearance.

   (b) It was the makeup that the performer who arrived very late applied before the stage appearance.

   Did the performer apply the makeup before the stage appearance?,Y

4. x-x-x dad etc pray dear mom therefore pre yours guy agree cancel worry grow Compare wisely.

   (a) It was the gun that the character who began the fight seized after the Western sunset.

   (b) It was the gun that the character who began the fight shot after the Western sunset.

   Was it the film that the character began?, N

5. x-x-x hell dad announces seem etc includes pre sexism mid implemented doubt ballet soul god whom pope.

   (a) It was the cigarette that the criminal who robbed the electronics store sought in the dark cell.

   (b) It was the cigarette that the criminal who robbed the electronics store smoked in the dark cell.

   Was it the cell in which the criminal smoked the cigarette?,Y

6. x-x-x eye etc thinks glad sir unto mid southern miss feedback vinyl pre cent rates imagine.

   (a) It was the button that the maid who returned from vacation spotted in the early morning.

   (b) It was the button that the maid who returned from vacation sewed in the early morning.

   Was the button sewed in the evening?, N

7. x-x-x dear etc woke glad sir berths pre domain map miles birthplace hurdle effort link assume seamless.

   (a) It was the gum that the camper who forgot to bring sandwiches fetched during the forest campfire.

   (b) It was the gum that the camper who forgot to bring sandwiches chewed during the forest campfire.

   Did the camper forget to bring sandwiches?,Y

8. x-x-x dear etc onto hear eat arrives mid epicenter yeah miss cent sorry fat aware glad.

   (a) It was the bill that the citizen who overslept for work sent after the phone call.

   (b) It was the bill that the citizen who overslept for work paid after the phone call.

   Did the citizen pay the bill before the phone call?, N

9. x-x-x dad etc pray seem lord toward fun strangers map posted launcher understood wish decide prevent.

(a) It was the solo that the singer who practiced for months dreaded throughout the entire evening.

(b) It was the solo that the singer who practiced for months performed throughout the entire evening.

Did the singer perform the solo?,Y

10. x-x-x hell mom swings want etc qui pre republic guys radius garter click hear protects lucky.

(a) It was the puzzle that the lad who received good grades mislaid after the mediocre lunch.

(b) It was the puzzle that the lad who received good grades solved after the mediocre lunch.

Was the lads lunch delicious?, N

11. x-x-x luck mom attends hear hate binaries yeah tyrants sir disagree voter review okay suppose answer.

(a) It was the perfume that the passerby who shopped for presents stole during the holiday season.

(b) It was the perfume that the passerby who shopped for presents smelled during the holiday season.

Did the passerby shop for presents?,Y

12. x-x-x dear kid else jobs sir anymore yeah radius miss wits disarray issues map noir aware.

(a) It was the hair that the student who rented a limo damaged before her prom night.

(b) It was the hair that the student who rented a limo shampooed before her prom night.

Was the limo rented by the teacher?, N

13. x-x-x dad etc nor grow lord anymore holy variety sir violations lakes guess feel recognise bureau.

(a) It was the car that the husband who changed his occupation washed after the delicious dinner.

(b) It was the car that the husband who changed his occupation drove after the delicious dinner .

Did the husband drive the car?,Y

14. x-x-x dad etc myself skin yeah regarding sick ourselves guy regardless crevices forget fat exist.

(a) It was the number that the secretary who collected the commission memorised during her break.

(b) It was the number that the secretary who collected the commission dialed during her break.

Did the secretary dial the number after her break?, N

15. x-x-x dad jobs pops wife cent avoid rationale yes bass rate includes yeah varieties jack etc aided Certified.

(a) It was the rug that the house attendant who quit his previous job destroyed on the rainy Wednesday.

(b) It was the rug that the house attendant who quit his previous job swept on the rainy Wednesday.

Was it the rug that was swept?,Y

16. x-x-x dear dad gotten mean jack osteoarthritis ball battering size Types Efforts emotion cent sir happen laws miss gathered.

(a) It was the museum that the philanthropist who journeyed to South America accessed in the middle of his vacation.

(b) It was the museum that the philanthropist who journeyed to South America founded in the middle of his vacation.

Did the philanthropist journey to Africa?, N

17. x-x-x dad cent elsewhere feet etc recommend lie trauma agree click monster jack yeah respectively.

(a) It was the newspaper that the assistant who stayed after hours removed from the headquarters.

(b) It was the newspaper that the assistant who stayed after hours printed from the headquarters.

Did the assistant print the newspaper?,Y

18. x-x-x hell kid else miss cent via mean reality minute happen cancel jack guys Regret eyes Park.

    (a) It was the food that the son who entered middle school dropped on the Fourth of July.

    (b) It was the food that the son who entered middle school cooked on the Fourth of July.

    Was it the daughter who entered middle school?, N

19. x-x-x eye dad adheres luck cent decide sale finish miss includes loyalties moment yeah moreover seriously.

    (a) It was the syringe that the doctor who opened a practice misused before the critical operation.

    (b) It was the syringe that the doctor who opened a practice sterilized before the critical operation.

    Was the syringe sterilized before the operation?,Y

20. x-x-x dad kid woke jobs cent preferring pre cardinal sun cops alleys jack yeah maybe worse.

    (a) It was the suit that the accountant who acquired a date rented at the local store.

    (b) It was the suit that the accountant who acquired a date ironed at the local store.

    Did the accountant iron the suit at his sisters store?, N

21. x-x-x mid mom exceeds dear etc via jobs portrait guys lady accept broader agency fund simplify respondents.

    (a) It was the whistle that the boy who listened to his mother damaged during the splendid celebration.

    (b) It was the whistle that the boy who listened to his mother blew during the splendid celebration.

    Did the boy listen to his mother?,Y

22. x-x-x hell etc eats know hate rely laws basis lord guy seem none treaties sad miss focuses damages.

(a) It was the pie that the chef who moved to a new town invented for the cooking contest.

(b) It was the pie that the chef who moved to a new town baked for the cooking contest.

Was it the chef who baked the pie?, N

23. x-x-x dad mom onto grow yeah extends sad duration miss responds feminists guys okay forth chronic.

(a) It was the card that the tourist who obtained the discount discarded by the movie theatre.

(b) It was the card that the tourist who obtained the discount activated by the movie theatre.

Was the card activated by the movie theatre?,Y

24. x-x-x dad sir behave hear etc degrade wild fiction soul learn sitcoms multi cent guys dumb.

(a) It was the jacket that the steward who crossed the world packed prior to the trip.

(b) It was the jacket that the steward who crossed the world unzipped prior to the trip.

Was the world crossed by the pilot?, N

25. x-x-x dad etc unless jobs mom depend duty reality kid yeah oxidation learn sit ourselves liability.

(a) It was the report that the writer who created the show misplaced after the afternoon broadcast.

(b) It was the report that the writer who created the show typed after the afternoon broadcast.

Was it the writer that typed the report?,Y

26. x-x-x dad mom survived hate dear occurring sick smelling pregnant weekend challenger ha guy carry thanksgiving.

(a) It was the hardware that the associate who mastered computer science exported at the large organization.

(b) It was the hardware that the associate who mastered computer science programmed at the large organization.

Did the associate program the software?, N

27. x-x-x hell dad else know etc thus pre approval target guys Global lack basis jack anywhere percent.

(a) It was the book that the lady who attended church last Sunday sold after the business meeting.

(b) It was the book that the lady who attended church last Sunday read after the business meeting.

Did the lady attend church last Sunday?,Y

28. x-x-x hell etc toward hurt sir alleging loss goal dad enhance schism basis hear enables economies.

(a) It was the orange that the traveler who lost his luggage rinsed after the delayed departure.

(b) It was the orange that the traveler who lost his luggage peeled after the delayed departure.

Was it the pear that the traveler peeled?, N

29. x-x-x hell dad exceed hurt risk whom mid remainder guys miss worse fascist worry web simplifies sexy.

(a) It was the flower that the wife who purchased the new house admired after the torrential rain.

(b) It was the flower that the wife who purchased the new house planted after the torrential rain.

Was the flower planted after the rain?,Y

30. x-x-x eye dad exited lord sir argued pre prince mark trophy fellows nation map yeah roman.

(a) It was the clutch that the driver who worked late nights repaired before the next shift.

(b) It was the clutch that the driver who worked late nights pressed before the next shift.

Did the driver press the brake?, N

31. x-x-x hell dad recover feet hear whom pre goal guy auxiliaries dynamics expect map glad chronic.

    (a) It was the alcohol that the girl who won the sweepstakes purchased before the rock concert.

    (b) It was the alcohol that the girl who won the sweepstakes poured before the rock concert.

    Did the girl win the sweepstakes?,Y

32. x-x-x eye dad exist cent hate anymore holy sweet term Commons rogue guess anti contributes slept.

    (a) It was the plane that the brother who moved to Arizona owned until the unfortunate crash.

    (b) It was the plane that the brother who moved to Arizona aviated until the unfortunate crash.

    Was it to Los Angelos that the brother moved?, N

33. x-x-x hell etc onto sick sad kilograms trip southern yeah paths studios okay mom derive daughters.

    (a) It was the code that the policeman who received a bonus cracked in the shabby apartment.

    (b) It was the code that the policeman who received a bonus entered in the shabby apartment.

    Was it the policeman who entered the code?,Y

34. x-x-x hell mom agree grow lord complicate cent congress guys combines delicacy al anti succeeds rights sell.

    (a) It was the radio that the adolescent who finished his homework played in the upstairs living room.

    (b) It was the radio that the adolescent who finished his homework unplugged in the upstairs living room.

    Was the radio unplugged in the bedroom?, N

35. x-x-x eye mom pray cent join entertain fun variation yeah worry backlash please map else besides.

(a) It was the drug that the physician who graduated last month misplaced before the next patient.

(b) It was the drug that the physician who graduated last month injected before the next patient.

Did the physician graduate last month?,Y

36. x-x-x hell dad pops ways drug forth yeah annual hazard thank non suite truth etc moreover avoid.

(a) It was the kite that the youth who turned twelve years old crashed after his birthday party.

(b) It was the kite that the youth who turned twelve years old flew after his birthday party.

Was the kite purchased by the youth?, N

5.3.3 Experiment 5 stimuli

Below are the experimental items from Experiment 5. Each item consists of the unrelated condition in (a), the related condition in (b) and the comprehension question followed by Y, if the correct answer was yes or N, if the correct answer was no.

1. (a) It was the boat that the guy who lived by the sea fixed in two sunny days.
   (b) It was the boat that the guy who lived by the sea sailed in two sunny days.

   Did the guy live by the sea?,Y

2. (a) It was the dress that the waitress who worked at the diner won on the last Monday of the month.
   (b) It was the dress that the waitress who worked at the diner wore on the last Monday of the month.

   Did the waitress wear the dress on Wednesday?, N

3. (a) It was the makeup that the performer who arrived very late ruined before the stage appearance.
   (b) It was the makeup that the performer who arrived very late applied before the stage appearance.

   Did the performer apply the makeup before the stage appearance?,Y

4. (a) It was the gun that the character who began the fight seized after the Western sunset.
   (b) It was the gun that the character who began the fight shot after the Western sunset.

   Was it the film that the character began?, N

5. (a) It was the cigarette that the criminal who robbed the electronics store sought in the dark cell.
   (b) It was the cigarette that the criminal who robbed the electronics store smoked in the dark cell.

   Was it the cell in which the criminal smoked the cigarette?,Y

6. (a) It was the button that the maid who returned from vacation spotted in the early morning.
   (b) It was the button that the maid who returned from vacation sewed in the early morning.

   Was the button sewed in the evening?, N

7. (a) It was the gum that the camper who forgot to bring sandwiches fetched during the forest campfire.

   (b) It was the gum that the camper who forgot to bring sandwiches chewed during the forest campfire.

   Did the camper forget to bring sandwiches?,Y

8. (a) It was the bill that the citizen who overslept for work sent after the phone call.

   (b) It was the bill that the citizen who overslept for work paid after the phone call.

   Did the citizen pay the bill before the phone call?, N

9. (a) It was the solo that the singer who practiced for months dreaded throughout the entire evening.

   (b) It was the solo that the singer who practiced for months performed throughout the entire evening.

   Did the singer perform the solo?,Y

10. (a) It was the puzzle that the lad who received good grades mislaid after the mediocre lunch.

    (b) It was the puzzle that the lad who received good grades solved after the mediocre lunch.

    Was the lads lunch delicious?, N

11. (a) It was the perfume that the passerby who shopped for presents stole during the holiday season.

    (b) It was the perfume that the passerby who shopped for presents smelled during the holiday season.

    Did the passerby shop for presents?,Y

12. (a) It was the hair that the student who rented a limo damaged before her prom night.

    (b) It was the hair that the student who rented a limo shampooed before her prom night.

    Was the limo rented by the teacher?, N

13. (a) It was the car that the husband who changed his occupation washed after the delicious dinner.

(b) It was the car that the husband who changed his occupation drove after the delicious dinner .

Did the husband drive the car?,Y

14. (a) It was the number that the secretary who collected the commission memorised during her break.

(b) It was the number that the secretary who collected the commission dialed during her break.

Did the secretary dial the number after her break?, N

15. (a) It was the rug that the house attendant who quit his previous job destroyed on the rainy Wednesday.

(b) It was the rug that the house attendant who quit his previous job swept on the rainy Wednesday.

Was it the rug that was swept?,Y

16. (a) It was the museum that the philanthropist who journeyed to South America accessed in the middle of his vacation.

(b) It was the museum that the philanthropist who journeyed to South America founded in the middle of his vacation.

Did the philanthropist journey to Africa?, N

17. (a) It was the newspaper that the assistant who stayed after hours removed from the headquarters.

(b) It was the newspaper that the assistant who stayed after hours printed from the headquarters.

Did the assistant print the newspaper?,Y

18. (a) It was the food that the son who entered middle school dropped on the Fourth of July.

(b) It was the food that the son who entered middle school cooked on the Fourth of July.

Was it the daughter who entered middle school?, N

19. (a) It was the syringe that the doctor who opened a practice misused before the critical operation.

(b) It was the syringe that the doctor who opened a practice sterilized before the critical operation.

Was the syringe sterilized before the operation?,Y

20. (a) It was the suit that the accountant who acquired a date rented at the local store.

    (b) It was the suit that the accountant who acquired a date ironed at the local store.

    Did the accountant iron the suit at his sisters store?, N

21. (a) It was the whistle that the boy who listened to his mother damaged during the splendid celebration.

    (b) It was the whistle that the boy who listened to his mother blew during the splendid celebration.

    Did the boy listen to his mother?,Y

22. (a) It was the pie that the chef who moved to a new town invented for the cooking contest.

    (b) It was the pie that the chef who moved to a new town baked for the cooking contest.

    Was it the chef who baked the pie?, N

23. (a) It was the card that the tourist who obtained the discount discarded by the movie theatre.

    (b) It was the card that the tourist who obtained the discount activated by the movie theatre.

    Was the card activated by the movie theatre?,Y

24. (a) It was the jacket that the steward who crossed the world packed prior to the trip.

    (b) It was the jacket that the steward who crossed the world unzipped prior to the trip.

    Was the world crossed by the pilot?, N

25. (a) It was the report that the writer who created the show misplaced after the afternoon broadcast.

    (b) It was the report that the writer who created the show typed after the afternoon broadcast.

    Was it the writer that typed the report?,Y

26. (a) It was the hardware that the associate who mastered computer science exported at the large organization.

(b) It was the hardware that the associate who mastered computer science programmed at the large organization.

Did the associate program the software?, N

27. (a) It was the book that the lady who attended church last Sunday sold after the business meeting.

(b) It was the book that the lady who attended church last Sunday read after the business meeting.

Did the lady attend church last Sunday?,Y

28. (a) It was the orange that the traveler who lost his luggage rinsed after the delayed departure.

(b) It was the orange that the traveler who lost his luggage peeled after the delayed departure.

Was it the pear that the traveler peeled?, N

29. (a) It was the flower that the wife who purchased the new house admired after the torrential rain.

(b) It was the flower that the wife who purchased the new house planted after the torrential rain.

Was the flower planted after the rain?,Y

30. (a) It was the clutch that the driver who worked late nights repaired before the next shift.

(b) It was the clutch that the driver who worked late nights pressed before the next shift.

Did the driver press the brake?, N

31. (a) It was the alcohol that the girl who won the sweepstakes purchased before the rock concert.

(b) It was the alcohol that the girl who won the sweepstakes poured before the rock concert.

Did the girl win the sweepstakes?,Y

32. (a) It was the plane that the brother who moved to Arizona owned until the unfortunate crash.

(b) It was the plane that the brother who moved to Arizona aviated until the unfortunate crash.

Was it to Los Angelos that the brother moved?, N

33. (a) It was the code that the policeman who received a bonus cracked in the shabby apartment.

    (b) It was the code that the policeman who received a bonus entered in the shabby apartment.

    Was it the policeman who entered the code?,Y

34. (a) It was the radio that the adolescent who finished his homework played in the upstairs living room.

    (b) It was the radio that the adolescent who finished his homework unplugged in the upstairs living room.

    Was the radio unplugged in the bedroom?, N

35. (a) It was the drug that the physician who graduated last month misplaced before the next patient.

    (b) It was the drug that the physician who graduated last month injected before the next patient.

    Did the physician graduate last month?,Y

36. (a) It was the kite that the youth who turned twelve years old crashed after his birthday party.

    (b) It was the kite that the youth who turned twelve years old flew after his birthday party.

    Was the kite purchased by the youth?, N

5.3.4 Experiment 7 stimuli

Below are the experimental items from Experiment 7. Each item consists of the plausible distractor/unambiguous verb condition in (a), the plausible distractor/ambiguous verb condition in (b), the implausible distractor/unambiguous verb condition in (c), the implausible distractor/ambiguous verb condition in (d), and the generated foil in (e).

1. (a) Sue remembered the letter that the butler with the cup had accidently shattered today in the dining room.

   (b) Sue remembered the letter that the butler with the cup accidently shattered today in the dining room.

   (c) Sue remembered the letter that the butler with the tie had accidently shattered today in the dining room.

   (d) Sue remembered the letter that the butler with the tie accidently shattered today in the dining room.

   (e) x-x-x propaganda etc afford lord dad impose goal grow nor self vegetative economies stars al guys namely duty.

2. (a) John noticed the tree that the camper by the chair had quietly moved before going off to sleep.

   (b) John noticed the tree that the camper by the chair quietly moved before going off to sleep.

   (c) John noticed the tree that the camper by the river had quietly moved before going off to sleep.

   (d) John noticed the tree that the camper by the river quietly moved before going off to sleep.

   (e) x-x-x outcome jack pays else cent hereby mid yeah yours pre treason clock career learn mom guys types.

3. (a) Jane saw the tent that the lady by the flowers had thoughtfully planted some years earlier.

   (b) Jane saw the tent that the lady by the flowers thoughtfully planted some years earlier.

   (c) Jane saw the tent that the lady by the fence had thoughtfully planted some years earlier.

   (d) Jane saw the tent that the lady by the fence thoughtfully planted some years earlier.

(e) x-x-x dear etc owns yeah cent onto pre anti aboard resignations handful soul click biggest.

4. (a) Jim saw the cutlery that the chef by the food had expertly cooked despite all of the pressure.

   (b) Jim saw the cutlery that the chef by the food expertly cooked despite all of the pressure.

   (c) Jim saw the cutlery that the chef by the sink had expertly cooked despite all of the pressure.

   (d) Jim saw the cutlery that the chef by the sink expertly cooked despite all of the pressure.

   (e) x-x-x cent glad alleges yeah nor vary holy sir deny plateaus chairs benefit hall fund vote estimate.

5. (a) Vicky noticed the gate that the farmer by the crops had recently harvested with a big new tractor.

   (b) Vicky noticed the gate that the farmer by the crops recently harvested with a big new tractor.

   (c) Vicky noticed the gate that the farmer by the fence had recently harvested with a big new tractor.

   (d) Vicky noticed the gate that the farmer by the fence recently harvested with a big new tractor.

   (e) x-x-x speaker eat rely yeah etc advise non pre renew politics countdown anti am guys else clinics.

6. (a) Lio liked the lake that the assistant with the present had beautifully wrapped with blue paper.

   (b) Lio liked the lake that the assistant with the present beautifully wrapped with blue paper.

   (c) Lio liked the lake that the assistant with the moustache had beautifully wrapped with blue paper.

   (d) Lio liked the lake that the assistant with the moustache beautifully wrapped with blue paper.

   (e) x-x-x yours cup asks glad holy ourselves hate ago feminism yeah imperialism offence rise earn worse.

7. (a) Gregory remembered the house that the client with the money had carefully deposited late in the evening.

(b) Gregory remembered the house that the client with the money carefully deposited late in the evening.

(c) Gregory remembered the house that the client with the cough had carefully deposited late in the evening.

(d) Gregory remembered the house that the client with the cough carefully deposited late in the evening.

(e) x-x-x governance jack shall yeah dad toward anti pre arise ambulance incumbent miss easy guys appears.

8. (a) Jessica liked the pen that the student with the script had passionately recited four hours ago.

(b) Jessica liked the pen that the student with the script passionately recited four hours ago.

(c) Jessica liked the pen that the student with the pencil had passionately recited four hours ago.

(d) Jessica liked the pen that the student with the pencil passionately recited four hours ago.

(e) x-x-x worse sir sec lose jack whether cool god varies sea thunderstorm rebates cent exist heat.

9. (a) Matthew remembered the pond that the man by the tree had expertly chopped to build a fire.

(b) Matthew remembered the pond that the man by the tree expertly chopped to build a fire.

(c) Matthew remembered the pond that the man by the lake had expertly chopped to build a fire.

(d) Matthew remembered the pond that the man by the lake expertly chopped to build a fire.

(e) x-x-x capitalism jack owns yeah mom nor al pre porn sir calipers mailing glad topic vote quit.

10. (a) Alice enjoyed the gift that the author with the story had vividly narrated at the school celebration yesterday.

(b) Alice enjoyed the gift that the author with the story vividly narrated at the school celebration yesterday.

(c) Alice enjoyed the gift that the author with the beard had vividly narrated at the school celebration yesterday.

(d) Alice enjoyed the gift that the author with the beard vividly narrated at the school celebration yesterday.

(e) x-x-x ceiling guys onto miss cent abroad laws etc awake sir bribery browsers glad met myself republicans reduction.

11. (a) Kate recalled the beach that her brother by the stairs had enthusiastically climbed a few hours ago.

(b) Kate recalled the beach that her brother by the stairs enthusiastically climbed a few hours ago.

(c) Kate recalled the beach that her brother by the pool had enthusiastically climbed a few hours ago.

(d) Kate recalled the beach that her brother by the pool enthusiastically climbed a few hours ago.

(e) x-x-x altitude cent yours else etc suppose fact anti submit non standardization nominee sir nor spend pre.

12. (a) Nigel saw the pavement that the guy near the rose had happily watered as the sun shone brightly.

(b) Nigel saw the pavement that the guy near the rose happily watered as the sun shone brightly.

(c) Nigel saw the pavement that the guy near the door had happily watered as the sun shone brightly.

(d) Nigel saw the pavement that the guy near the door happily watered as the sun shone brightly.

(e) x-x-x wall sir squamous dear pre nor holy laws earn vaccine wartime glad map via tenet bottling.

13. (a) Monica saw the building that the man by the cart had hastily pushed while looking for food.

(b) Monica saw the building that the man by the cart hastily pushed while looking for food.

(c) Monica saw the building that the man by the shop had hastily pushed while looking for food.

(d) Monica saw the building that the man by the shop hastily pushed while looking for food.

(e) x-x-x pre glad depends yeah etc whom al sir unto nephews hippos prime variety wear cent.

14. (a) Julia saw the beer that the lady with the meal had happily chewed during an expensive night out.

    (b) Julia saw the beer that the lady with the meal happily chewed during an expensive night out.

    (c) Julia saw the beer that the lady with the wine had happily chewed during an expensive night out.

    (d) Julia saw the beer that the lady with the wine happily chewed during an expensive night out.

    (e) x-x-x sale etc helps dear sir thus data bus info yeah handful pagans mental guys solutions goals yes.

15. (a) Warren noticed the shirt that the girl by the tiles had accidently cracked earlier near the sink.

    (b) Warren noticed the shirt that the girl by the tiles accidently cracked earlier near the sink.

    (c) Warren noticed the shirt that the girl by the food had accidently cracked earlier near the sink.

    (d) Warren noticed the shirt that the girl by the food accidently cracked earlier near the sink.

    (e) x-x-x citizen cent yours yeah etc thus mid till euros okay conspiring trilogy awesome goal lord unto.

16. (a) Hillary remembered the glass that the teenager with the hat had happily sported during a big night out.

    (b) Hillary remembered the glass that the teenager with the hat happily sported during a big night out.

    (c) Hillary remembered the glass that the teenager with the pen had happily sported during a big night out.

    (d) Hillary remembered the glass that the teenager with the pen happily sported during a big night out.

    (e) x-x-x unexpected cent aware hope etc likewise anti yeah rely ways valleys earthen minute guys ones terms laws.

17. (a) Barry saw the potato that the housewife with the ornament had carefully dusted two hours ago.

    (b) Barry saw the potato that the housewife with the ornament carefully dusted two hours ago.

(c) Barry saw the potato that the housewife with the shopping had carefully dusted two hours ago.

(d) Barry saw the potato that the housewife with the shopping carefully dusted two hours ago.

(e) x-x-x hell etc seeks glad nor formulate okay hear stadiums sir coalition cranes god shall non.

18. (a) Joanna noticed the spoon that the boy with the juice had stupidly poured during a break at school.

(b) Joanna noticed the spoon that the boy with the juice stupidly poured during a break at school.

(c) Joanna noticed the spoon that the boy with the napkin had stupidly poured during a break at school.

(d) Joanna noticed the spoon that the boy with the napkin stupidly poured during a break at school.

(e) x-x-x citizen cent relax hate etc nor laws fund deems ordnance capita nation idea users kids answer.

19. (a) Nathan noticed the bag that the student with the essay had patiently drafted during a dull class.

(b) Nathan noticed the bag that the student with the essay patiently drafted during a dull class.

(c) Nathan noticed the bag that the student with the ruler had patiently drafted during a dull class.

(d) Nathan noticed the bag that the student with the ruler patiently drafted during a dull class.

(e) x-x-x routine cent nor anti pre anymore ways yeah alike amplitude mountain wonder le fats teams.

20. (a) Sam found the earring that his grandmother with the blanket had diligently knitted during last winter.

(b) Sam found the earring that his grandmother with the blanket diligently knitted during last winter.

(c) Sam found the earring that his grandmother with the lipstick had diligently knitted during last winter.

(d) Sam found the earring that his grandmother with the lipstick diligently knitted during last winter.

(e) x-x-x doubt etc likened yeah pre accountable anti map secondly prefecture harpist nation hear senate.

21. (a) Alex cherished the chair that his son with the poem had artfully composed during the holidays last summer.

(b) Alex cherished the chair that his son with the poem artfully composed during the holidays last summer.

(c) Alex cherished the chair that his son with the car had artfully composed during the holidays last summer.

(d) Alex cherished the chair that his son with the car artfully composed during the holidays last summer.

(e) x-x-x harmonies cent onto miss etc pre trip map nor positron democrat forget lady analyzed guys bottom.

22. (a) Annabelle saw the knife that the cook with the pork had expertly grilled during the rush at lunch.

(b) Annabelle saw the knife that the cook with the pork expertly grilled during the rush at lunch.

(c) Annabelle saw the knife that the cook with the fork had expertly grilled during the rush at lunch.

(d) Annabelle saw the knife that the cook with the fork expertly grilled during the rush at lunch.

(e) x-x-x okay yeah cared anti mom onto goal fund laps tsunamis mayoral career hear owns mind loans.

23. (a) Jake noticed the automobile that the pilot by the plane had expertly landed a couple of nights ago.

(b) Jake noticed the automobile that the pilot by the plane expertly landed a couple of nights ago.

(c) Jake noticed the automobile that the pilot by the tank had expertly landed a couple of nights ago.

(d) Jake noticed the automobile that the pilot by the tank expertly landed a couple of nights ago.

(e) x-x-x biology hall facilitate cent etc relax pre till trans carbonyl window fact itself laws brings lord.

24. (a) Wendy remembered the beach that the man with the boat had confidently sailed one hot and sunny day.

(b) Wendy remembered the beach that the man with the boat confidently sailed one hot and sunny day.

(c) Wendy remembered the beach that the man with the shirt had confidently sailed one hot and sunny day.

(d) Wendy remembered the beach that the man with the shirt confidently sailed one hot and sunny day.

(e) x-x-x indication jack watts mind nor till ones map onto mid polytechnic gallon soul laws pre wills sir.

25. (a) Barry saw the stage that the pupil by the guitar had passionately played during a recital one day.

(b) Barry saw the stage that the pupil by the guitar passionately played during a recital one day.

(c) Barry saw the stage that the pupil by the curtain had passionately played during a recital one day.

(d) Barry saw the stage that the pupil by the curtain passionately played during a recital one day.

(e) x-x-x goal cent meant dear sir avail al anti agrees wild biodiversity prince forget guys grasses want risk.

26. (a) Jessica saw the handbag that the lady by the bread had carefully baked while in the kitchen today.

(b) Jessica saw the handbag that the lady by the bread carefully baked while in the kitchen today.

(c) Jessica saw the handbag that the lady by the kettle had carefully baked while in the kitchen today.

(d) Jessica saw the handbag that the lady by the kettle carefully baked while in the kitchen today.

(e) x-x-x jack cent briefed yeah pre else mid till fined activists desks click term glad reached weird.

27. (a) Bob noticed the towel that the boy with the apple had slowly peeled after relaxing in the sun.

(b) Bob noticed the towel that the boy with the apple slowly peeled after relaxing in the sun.

(c) Bob noticed the towel that the boy with the juice had slowly peeled after relaxing in the sun.

(d) Bob noticed the towel that the boy with the juice slowly peeled after relaxing in the sun.

(e) x-x-x seconds cent donate anti sir nor holy yeah firms fiscal racket super fortress map link whom.

28. (a) Janice admired the fireplace that the woman with the scarf had happily knitted while at home last week.

(b) Janice admired the fireplace that the woman with the scarf happily knitted while at home last week.

(c) Janice admired the fireplace that the woman with the paper had happily knitted while at home last week.

(d) Janice admired the fireplace that the woman with the paper happily knitted while at home last week.

(e) x-x-x voltage door publishes anti cent shall holy map renew handful better click file laws link hate.

29. (a) Melissa liked the teapot that the woman with the bread had quickly toasted one busy day before work.

(b) Melissa liked the teapot that the woman with the bread quickly toasted one busy day before work.

(c) Melissa liked the teapot that the woman with the plate had quickly toasted one busy day before work.

(d) Melissa liked the teapot that the woman with the plate quickly toasted one busy day before work.

(e) x-x-x worse sir livres hear holy shall jack etc forth premier cartels pre wars glad rights cent.

30. (a) Charlie liked the glass that the chef with the fish had patiently cooked in the restaurants kitchen.

(b) Charlie liked the glass that the chef with the fish patiently cooked in the restaurants kitchen.

(c) Charlie liked the glass that the chef with the apron had patiently cooked in the restaurants kitchen.

(d) Charlie liked the glass that the chef with the apron patiently cooked in the restaurants kitchen.

(e) x-x-x loans link euros anti etc unto holy lord amid universes balloon fund miss resignations ignored.

31. (a) Sophia liked the music that the girl with the coat had happily sported to the party last weekend.

    (b) Sophia liked the music that the girl with the coat happily sported to the party last weekend.

    (c) Sophia liked the music that the girl with the money had happily sported to the party last weekend.

    (d) Sophia liked the music that the girl with the money happily sported to the party last weekend.

    (e) x-x-x loose hour shall miss cent thus hell via occur yeah handful couplet link guys aware glad experts.

32. (a) Vince found the bandage that the soldier near the gun had carelessly fired before taking time to aim.

    (b) Vince found the bandage that the soldier near the gun carelessly fired before taking time to aim.

    (c) Vince found the bandage that the soldier near the house had carelessly fired before taking time to aim.

    (d) Vince found the bandage that the soldier near the house carelessly fired before taking time to aim.

    (e) x-x-x whole jack thereof yeah than whereas lips laws nor symbolises roman signed region wish soul sins.

33. (a) Emma heard the bird that the child with the flute had skillfully played during the show yesterday.

    (b) Emma heard the bird that the child with the flute skillfully played during the show yesterday.

    (c) Emma heard the bird that the child with the hat had skillfully played during the show yesterday.

    (d) Emma heard the bird that the child with the hat skillfully played during the show yesterday.

    (e) x-x-x yours etc ours anti sir twice ways fund grabs goal nonfiction growth cities guys goes solutions.

34. (a) Elizabeth smelled the perfume that the lady with the cake had freshly baked in the bakery this morning.

    (b) Elizabeth smelled the perfume that the lady with the cake freshly baked in the bakery this morning.

(c) Elizabeth smelled the perfume that the lady with the earrings had freshly baked in the bakery this morning.

(d) Elizabeth smelled the perfume that the lady with the earrings freshly baked in the bakery this morning.

(e) x-x-x remorse kid grabbed yeah etc thus data map analysts lawsuit pools loss guys levied miss percent.

35. (a) Rose saw the chair that the waiter with the coffee had carefully poured while serving breakfast.

(b) Rose saw the chair that the waiter with the coffee carefully poured while serving breakfast.

(c) Rose saw the chair that the waiter with the tray had carefully poured while serving breakfast.

(d) Rose saw the chair that the waiter with the tray carefully poured while serving breakfast.

(e) x-x-x sort etc owned yeah sir expire east cent argues intensity ranger prime pension languages.

36. (a) Jennifer admired the hat that the woman with the car had cleverly parked by the yellow house.

(b) Jennifer admired the hat that the woman with the car cleverly parked by the yellow house.

(c) Jennifer admired the hat that the woman with the suitcase had cleverly parked by the yellow house.

(d) Jennifer admired the hat that the woman with the suitcase cleverly parked by the yellow house.

(e) x-x-x sectors door rotated holy east shall hell map follows emeritus intern hope pre belong worry.

37. (a) Thomas smelled the window that the neighbour with the lily had patiently watered earlier in the morning.

(b) Thomas smelled the window that the neighbour with the lily patiently watered earlier in the morning.

(c) Thomas smelled the window that the neighbour with the briefcase had patiently watered earlier in the morning.

(d) Thomas smelled the window that the neighbour with the briefcase patiently watered earlier in the morning.

(e) x-x-x addicts sir yours yeah dad vacations goal etc primetime insurgent tallest sources okay know realize.

38. (a) Harry noticed the lamp that the cook with the carrot had quickly chopped on the counter an hour ago.

(b) Harry noticed the lamp that the cook with the carrot quickly chopped on the counter an hour ago.

(c) Harry noticed the lamp that the cook with the shirt had quickly chopped on the counter an hour ago.

(d) Harry noticed the lamp that the cook with the shirt quickly chopped on the counter an hour ago.

(e) x-x-x outcome bill sings hate etc rely yeah vote renew anti weekend militia risk lord beliefs al mode laws.

39. (a) Daniel admired the pond that the farmer by the rye had efficiently harvested on a warm sunny day.

(b) Daniel admired the pond that the farmer by the rye efficiently harvested on a warm sunny day.

(c) Daniel admired the pond that the farmer by the garage had efficiently harvested on a warm sunny day.

(d) Daniel admired the pond that the farmer by the garage efficiently harvested on a warm sunny day.

(e) x-x-x throats miss unto hear sir advise al yeah facto secretariat travelers god glad seek firms send.

40. (a) Helen remembered the present that the couple by the willow had carefully planted a decade ago.

(b) Helen remembered the present that the couple by the willow carefully planted a decade ago.

(c) Helen remembered the present that the couple by the river had carefully planted a decade ago.

(d) Helen remembered the present that the couple by the river carefully planted a decade ago.

(e) x-x-x capitalist etc towards shut pre anyway sir fun opted diversity essence hate equity dad.

5.3.5 Experiment 8 stimuli

Below are the experimental items from Experiment 8. Each item consists of the plausible distractor condition in (a), the implausible distractor condition in (b), and the generated foil in (c).

1. (a) Sue remembered the letter that the butler with the cup had recently written.
   (b) Sue remembered the letter that the butler with the tie had recently written.
   (c) x-x-x propaganda etc afford lord dad impose goal grow nor self vegetative shattered.

2. (a) John noticed the tree that the camper by the chair had carefully planted.
   (b) John noticed the tree that the camper by the river had carefully planted.
   (c) x-x-x outcome jack pays else cent hereby mid yeah yours pre treason moved.

3. (a) Jane saw the tent that the lady by the flowers had thoughtfully folded.
   (b) Jane saw the tent that the lady by the fence had thoughtfully folded.
   (c) x-x-x dear etc owns yeah cent onto pre anti aboard hell resignations planted.

4. (a) Jim saw the cutlery that the chef by the food had patiently sharpened.
   (b) Jim saw the cutlery that the chef by the sink had patiently sharpened.
   (c) x-x-x cent glad alleges yeah nor vary holy sir deny pre plateaus cooked.

5. (a) Vicky noticed the gate that the farmer by the crops had recently locked.
   (b) Vicky noticed the gate that the farmer by the fence had recently locked.
   (c) x-x-x speaker eat rely yeah etc advise non pre renew soul politics harvested.

6. (a) Lio liked the lake that the assistant with the present had carefully crossed.
   (b) Lio liked the lake that the assistant with the moustache had carefully crossed.
   (c) x-x-x yours cup asks glad holy ourselves hate ago feminism yeah imperialism wrapped.

7. (a) Gregory remembered the house that the client with the money had recently built.
   (b) Gregory remembered the house that the client with the cough had recently built.
   (c) x-x-x governance jack shall yeah dad toward anti pre arise sir ambulance deposited.

8. (a) Jessica liked the pen that the student with the script had quickly refilled.
   (b) Jessica liked the pen that the student with the pencil had quickly refilled.
   (c) x-x-x worse sir sec lose jack whether cool god varies sea thunderstorm recited.

9. (a) Matthew remembered the pond that the man by the tree had enthusiastically built.
   (b) Matthew remembered the pond that the man by the lake had enthusiastically built.
   (c) x-x-x capitalism jack owns yeah mom nor al pre porn sir calipers chopped.

10. (a) Alice enjoyed the gift that the author with the story had happily given.
    (b) Alice enjoyed the gift that the author with the beard had happily given.
    (c) x-x-x ceiling guys onto miss cent abroad laws etc awake sir bribery narrated.

11. (a) Kate recalled the beach that her brother by the stairs had keenly explored.
    (b) Kate recalled the beach that her brother by the pool had keenly explored.
    (c) x-x-x altitude cent yours else etc suppose fact anti submit non standardization climbed.

12. (a) Nigel saw the pavement that the guy near the rose had happily walked.
    (b) Nigel saw the pavement that the guy near the door had happily walked.
    (c) x-x-x wall sir squamous dear pre nor holy laws earn anti vaccine watered.

13. (a) Monica saw the building that the man by the cart had hastily engineered.
    (b) Monica saw the building that the man by the shop had hastily engineered.
    (c) x-x-x pre glad depends yeah etc whom al sir unto hate nephews pushed.

14. (a) Julia saw the beer that the lady with the meal had happily brewed.
    (b) Julia saw the beer that the lady with the wine had happily brewed.
    (c) x-x-x sale etc helps dear sir thus data bus info yeah handful chewed.

15. (a) Warren noticed the shirt that the girl by the tiles had accidently worn.
    (b) Warren noticed the shirt that the girl by the food had accidently worn.
    (c) x-x-x citizen cent yours yeah etc thus mid till euros okay conspiring cracked.

16. (a) Hillary remembered the glass that the teenager with the hat had previously filled.

(b) Hillary remembered the glass that the teenager with the pen had previously filled.

(c) x-x-x unexpected cent aware hope etc likewise anti yeah rely ways valleys sported.

17. (a) Barry saw the potato that the housewife with the ornament had carefully sliced.

(b) Barry saw the potato that the housewife with the shopping had carefully sliced.

(c) x-x-x hell etc seeks glad nor formulate okay hear stadiums sir coalition dusted.

18. (a) Joanna noticed the spoon that the boy with the juice had clumsily bent.

(b) Joanna noticed the spoon that the boy with the napkin had clumsily bent.

(c) x-x-x citizen cent relax hate etc nor laws fund deems yeah ordnance poured.

19. (a) Nathan noticed the bag that the student with the essay had hastily packed.

(b) Nathan noticed the bag that the student with the ruler had hastily packed.

(c) x-x-x routine cent nor anti pre anymore ways yeah alike mid amplitude drafted.

20. (a) Sam found the earrings that his grandmother with the blanket had recently polished.

(b) Sam found the earrings that his grandmother with the lipstick had recently polished.

(c) x-x-x doubt etc likened yeah pre accountable anti map secondly sir prefecture knitted.

21. (a) Alex cherished the chair that his son with the poem had artfully carved.

(b) Alex cherished the chair that his son with the car had artfully carved.

(c) x-x-x harmonies cent onto miss etc pre trip map nor anti positron composed.

22. (a) Annabelle saw the knife that the cook with the pork had expertly sharpened.

(b) Annabelle saw the knife that the cook with the fork had expertly sharpened.

(c) x-x-x okay yeah cared anti mom onto goal fund laps hell tsunamis grilled.

23. (a) Jake noticed the car that the pilot by the plane had previously borrowed.
    (b) Jake noticed the car that the pilot by the tank had previously borrowed.
    (c) x-x-x biology hall facilitate cent etc relax pre till trans yeah carbonyl landed.

24. (a) Wendy remembered the beach that the man with the boat had confidently patrolled.
    (b) Wendy remembered the beach that the man with the shirt had confidently patrolled.
    (c) x-x-x indication jack watts mind nor till ones map onto mid polytechnic sailed.

25. (a) Barry saw the stage that the pupil by the guitar had unexpectedly mopped.
    (b) Barry saw the stage that the pupil by the curtain had unexpectedly mopped.
    (c) x-x-x goal cent meant dear sir avail al anti agrees wild biodiversity played.

26. (a) Jessica saw the handbag that the lady by the bread had carefully crocheted.
    (b) Jessica saw the handbag that the lady by the kettle had carefully crocheted.
    (c) x-x-x jack cent briefed yeah pre else mid till fined sir activists baked.

27. (a) Bob noticed the towel that the boy with the apple had slowly dried.
    (b) Bob noticed the towel that the boy with the juice had slowly dried.
    (c) x-x-x seconds cent donate anti sir nor holy yeah firms pre fiscal peeled.

28. (a) Janice admired the fireplace that the woman with the scarf had happily built.
    (b) Janice admired the fireplace that the woman with the paper had happily built.
    (c) x-x-x voltage door publishes anti cent shall holy map renew hell handful knitted.

29. (a) Melissa liked the teapot that the woman with the bread had quickly poured.
    (b) Melissa liked the teapot that the woman with the plate had quickly poured.
    (c) x-x-x worse sir livres hear holy shall jack etc forth yeah premier toasted.

30. (a) Charlie liked the glass that the chef with the fish had sloppily broken.
    (b) Charlie liked the glass that the chef with the apron had sloppily broken.
    (c) x-x-x loans link euros anti etc unto holy lord amid non universes cooked.

31. (a) Sophia liked the music that the girl with the coat had happily played.

(b) Sophia liked the music that the girl with the money had happily played.

(c) x-x-x loose hour shall miss cent thus hell via occur yeah handful sported.

32. (a) Vince found the bandage that the soldier near the gun had speedily applied.

(b) Vince found the bandage that the soldier near the house had speedily applied.

(c) x-x-x whole jack thereof yeah than whereas lips laws nor pre symbolises fired.

33. (a) Emma heard the bird that the child with the flute had patiently watched.

(b) Emma heard the bird that the child with the hat had patiently watched.

(c) x-x-x yours etc ours anti sir twice ways fund grabs goal nonfiction played.

34. (a) Elizabeth smelled the perfume that the lady with the cake had always sprayed.

(b) Elizabeth smelled the perfume that the lady with the earrings had always sprayed.

(c) x-x-x remorse kid grabbed yeah etc thus data map analysts sir lawsuit baked.

35. (a) Rose saw the chair that the waiter with the coffee had carefully reupholstered.

(b) Rose saw the chair that the waiter with the tray had carefully reupholstered.

(c) x-x-x sort etc owned yeah sir expire east cent argues hope intensity poured.

36. (a) Jennifer admired the hat that the woman with the car had cleverly worn.

(b) Jennifer admired the hat that the woman with the suitcase had cleverly worn.

(c) x-x-x sectors door rotated holy east shall hell map follows sir emeritus parked.

37. (a) Thomas saw the window that the neighbour with the lily had neglectfully smashed.

(b) Thomas saw the window that the neighbour with the briefcase had neglectfully smashed.

(c) x-x-x addicts sir yours yeah dad vacations goal etc primetime pre insurgent watered.

38. (a) Harry noticed the torch that the cook with the carrot had quickly lit.

    (b) Harry noticed the torch that the cook with the shirt had quickly lit.

    (c) x-x-x outcome bill sings hate etc rely yeah vote renew anti weekend chopped.

39. (a) Daniel admired the pond that the farmer by the rye had efficiently drained.

    (b) Daniel admired the pond that the farmer by the garage had efficiently drained.

    (c) x-x-x throats miss unto hear sir advise al yeah facto pre secretariat harvested.

40. (a) Helen remembered the present that the couple by the willow had carefully picked.

    (b) Helen remembered the present that the couple by the river had carefully picked.

    (c) x-x-x capitalist etc towards shut pre anyway sir fun opted cent diversity planted.

5.3.6 Experiment 9 stimuli

Below are the experimental items from Experiment 9. Each item consists of the plausible sentence/plausible distractor condition in (a), the plausible sentence/implausible distractor condition in (b), the implausible sentence/plausible distractor condition in (c), and the implausible sentence/implausible distractor condition in (d).

1. (a) Sue remembered the letter that the butler with the cup had recently written.
   (b) Sue remembered the letter that the butler with the cup had recently shattered.
   (c) Sue remembered the letter that the butler with the tie had recently written.
   (d) Sue remembered the letter that the butler with the tie had recently shattered.

2. (a) John noticed the tree that the camper by the chair had carefully planted.
   (b) John noticed the tree that the camper by the chair had carefully moved.
   (c) John noticed the tree that the camper by the river had carefully planted.
   (d) John noticed the tree that the camper by the river had carefully moved.

3. (a) Jane saw the tent that the lady by the flowers had thoughtfully folded.
   (b) Jane saw the tent that the lady by the flowers had thoughtfully planted.
   (c) Jane saw the tent that the lady by the fence had thoughtfully folded.
   (d) Jane saw the tent that the lady by the fence had thoughtfully planted.

4. (a) Jim saw the cutlery that the chef by the food had patiently sharpened.
   (b) Jim saw the cutlery that the chef by the food had patiently cooked.
   (c) Jim saw the cutlery that the chef by the sink had patiently sharpened.
   (d) Jim saw the cutlery that the chef by the sink had patiently cooked.

5. (a) Vicky noticed the gate that the farmer by the crops had recently locked.
   (b) Vicky noticed the gate that the farmer by the crops had recently harvested.
   (c) Vicky noticed the gate that the farmer by the fence had recently locked.
   (d) Vicky noticed the gate that the farmer by the fence had recently harvested.

6. (a) Lio liked the lake that the assistant with the present had carefully crossed.
   (b) Lio liked the lake that the assistant with the present had carefully wrapped.
   (c) Lio liked the lake that the assistant with the moustache had carefully crossed.

(d) Lio liked the lake that the assistant with the moustache had carefully wrapped.

7. (a) Gregory remembered the house that the client with the money had recently built.

   (b) Gregory remembered the house that the client with the money had recently deposited.

   (c) Gregory remembered the house that the client with the cough had recently built.

   (d) Gregory remembered the house that the client with the cough had recently deposited.

8. (a) Jessica liked the pen that the student with the script had quickly refilled.

   (b) Jessica liked the pen that the student with the script had quickly recited.

   (c) Jessica liked the pen that the student with the pencil had quickly refilled.

   (d) Jessica liked the pen that the student with the pencil had quickly recited.

9. (a) Matthew remembered the pond that the man by the tree had enthusiastically built.

   (b) Matthew remembered the pond that the man by the tree had enthusiastically chopped.

   (c) Matthew remembered the pond that the man by the lake had enthusiastically built.

   (d) Matthew remembered the pond that the man by the lake had enthusiastically chopped.

10. (a) Alice enjoyed the gift that the author with the story had happily given.

    (b) Alice enjoyed the gift that the author with the story had happily narrated.

    (c) Alice enjoyed the gift that the author with the beard had happily given.

    (d) Alice enjoyed the gift that the author with the beard had happily narrated.

11. (a) Kate recalled the beach that her brother by the stairs had keenly explored.

    (b) Kate recalled the beach that her brother by the stairs had keenly climbed.

    (c) Kate recalled the beach that her brother by the pool had keenly explored.

    (d) Kate recalled the beach that her brother by the pool had keenly climbed.

12. (a) Nigel saw the pavement that the guy near the rose had happily walked.

    (b) Nigel saw the pavement that the guy near the rose had happily watered.

(c) Nigel saw the pavement that the guy near the door had happily walked.

(d) Nigel saw the pavement that the guy near the door had happily watered.

13. (a) Monica saw the building that the man by the cart had hastily engineered.

(b) Monica saw the building that the man by the cart had hastily pushed.

(c) Monica saw the building that the man by the shop had hastily engineered.

(d) Monica saw the building that the man by the shop had hastily pushed.

14. (a) Julia saw the beer that the lady with the meal had happily brewed.

(b) Julia saw the beer that the lady with the meal had happily chewed.

(c) Julia saw the beer that the lady with the wine had happily brewed.

(d) Julia saw the beer that the lady with the wine had happily chewed.

15. (a) Warren noticed the shirt that the girl by the tiles had accidently worn.

(b) Warren noticed the shirt that the girl by the tiles had accidently cracked.

(c) Warren noticed the shirt that the girl by the food had accidently worn.

(d) Warren noticed the shirt that the girl by the food had accidently cracked.

16. (a) Hillary remembered the glass that the teenager with the hat had previously filled.

(b) Hillary remembered the glass that the teenager with the hat had previously sported.

(c) Hillary remembered the glass that the teenager with the pen had previously filled.

(d) Hillary remembered the glass that the teenager with the pen had previously sported.

17. (a) Barry saw the potato that the housewife with the ornament had carefully sliced.

(b) Barry saw the potato that the housewife with the ornament had carefully dusted.

(c) Barry saw the potato that the housewife with the shopping had carefully sliced.

(d) Barry saw the potato that the housewife with the shopping had carefully dusted.

18. (a) Joanna noticed the spoon that the boy with the juice had clumsily bent.

(b) Joanna noticed the spoon that the boy with the juice had clumsily poured.

(c) Joanna noticed the spoon that the boy with the napkin had clumsily bent.

(d) Joanna noticed the spoon that the boy with the napkin had clumsily poured.

19. (a) Nathan noticed the bag that the student with the essay had hastily packed.

(b) Nathan noticed the bag that the student with the essay had hastily drafted.

(c) Nathan noticed the bag that the student with the ruler had hastily packed.

(d) Nathan noticed the bag that the student with the ruler had hastily drafted.

20. (a) Sam found the earrings that his grandmother with the blanket had recently polished.

(b) Sam found the earrings that his grandmother with the blanket had recently knitted.

(c) Sam found the earrings that his grandmother with the lipstick had recently polished.

(d) Sam found the earrings that his grandmother with the lipstick had recently knitted.

21. (a) Alex cherished the chair that his son with the poem had artfully carved.

(b) Alex cherished the chair that his son with the poem had artfully composed.

(c) Alex cherished the chair that his son with the car had artfully carved.

(d) Alex cherished the chair that his son with the car had artfully composed.

22. (a) Annabelle saw the knife that the cook with the pork had expertly sharpened.

(b) Annabelle saw the knife that the cook with the pork had expertly grilled.

(c) Annabelle saw the knife that the cook with the fork had expertly sharpened.

(d) Annabelle saw the knife that the cook with the fork had expertly grilled.

23. (a) Jake noticed the car that the pilot by the plane had previously borrowed.

(b) Jake noticed the car that the pilot by the plane had previously landed.

(c) Jake noticed the car that the pilot by the tank had previously borrowed.

(d) Jake noticed the car that the pilot by the tank had previously landed.

24. (a) Wendy remembered the beach that the man with the boat had confidently patrolled.

(b) Wendy remembered the beach that the man with the boat had confidently sailed.

(c) Wendy remembered the beach that the man with the shirt had confidently patrolled.

(d) Wendy remembered the beach that the man with the shirt had confidently sailed.

25. (a) Barry saw the stage that the pupil by the guitar had unexpectedly mopped.

(b) Barry saw the stage that the pupil by the guitar had unexpectedly played.

(c) Barry saw the stage that the pupil by the curtain had unexpectedly mopped.

(d) Barry saw the stage that the pupil by the curtain had unexpectedly played.

26. (a) Jessica saw the handbag that the lady by the bread had carefully crocheted.

(b) Jessica saw the handbag that the lady by the bread had carefully baked.

(c) Jessica saw the handbag that the lady by the kettle had carefully crocheted.

(d) Jessica saw the handbag that the lady by the kettle had carefully baked.

27. (a) Bob noticed the towel that the boy with the apple had slowly dried.

(b) Bob noticed the towel that the boy with the apple had slowly peeled.

(c) Bob noticed the towel that the boy with the juice had slowly dried.

(d) Bob noticed the towel that the boy with the juice had slowly peeled.

28. (a) Janice admired the fireplace that the woman with the scarf had happily built.

(b) Janice admired the fireplace that the woman with the scarf had happily knitted.

(c) Janice admired the fireplace that the woman with the paper had happily built.

(d) Janice admired the fireplace that the woman with the paper had happily knitted.

29. (a) Melissa liked the teapot that the woman with the bread had quickly poured.

(b) Melissa liked the teapot that the woman with the bread had quickly toasted.

(c) Melissa liked the teapot that the woman with the plate had quickly poured.

(d) Melissa liked the teapot that the woman with the plate had quickly toasted.

30. (a) Charlie liked the glass that the chef with the fish had sloppily broken.

(b) Charlie liked the glass that the chef with the fish had sloppily cooked.

(c) Charlie liked the glass that the chef with the apron had sloppily broken.

(d) Charlie liked the glass that the chef with the apron had sloppily cooked.

31. (a) Sophia liked the music that the girl with the coat had happily played.

(b) Sophia liked the music that the girl with the coat had happily sported.

(c) Sophia liked the music that the girl with the money had happily played.

(d) Sophia liked the music that the girl with the money had happily sported.

32. (a) Vince found the bandage that the soldier near the gun had speedily applied.

(b) Vince found the bandage that the soldier near the gun had speedily fired.

(c) Vince found the bandage that the soldier near the house had speedily applied.

(d) Vince found the bandage that the soldier near the house had speedily fired.

33. (a) Emma heard the bird that the child with the flute had patiently watched.

(b) Emma heard the bird that the child with the flute had patiently played.

(c) Emma heard the bird that the child with the hat had patiently watched.

(d) Emma heard the bird that the child with the hat had patiently played.

34. (a) Elizabeth smelled the perfume that the lady with the cake had always sprayed.

(b) Elizabeth smelled the perfume that the lady with the cake had always baked.

(c) Elizabeth smelled the perfume that the lady with the earrings had always sprayed.

(d) Elizabeth smelled the perfume that the lady with the earrings had always baked.

35. (a) Rose saw the chair that the waiter with the coffee had carefully reupholstered.

(b) Rose saw the chair that the waiter with the coffee had carefully poured.

(c) Rose saw the chair that the waiter with the tray had carefully reupholstered.

(d) Rose saw the chair that the waiter with the tray had carefully poured.

36. (a) Jennifer admired the hat that the woman with the car had cleverly worn.
    (b) Jennifer admired the hat that the woman with the car had cleverly parked.
    (c) Jennifer admired the hat that the woman with the suitcase had cleverly worn.
    (d) Jennifer admired the hat that the woman with the suitcase had cleverly parked.

37. (a) Thomas saw the window that the neighbour with the lily had neglectfully smashed.
    (b) Thomas saw the window that the neighbour with the lily had neglectfully watered.
    (c) Thomas saw the window that the neighbour with the briefcase had neglectfully smashed.
    (d) Thomas saw the window that the neighbour with the briefcase had neglectfully watered.

38. (a) Harry noticed the torch that the cook with the carrot had quickly lit.
    (b) Harry noticed the torch that the cook with the carrot had quickly chopped.
    (c) Harry noticed the torch that the cook with the shirt had quickly lit.
    (d) Harry noticed the torch that the cook with the shirt had quickly chopped.

39. (a) Daniel admired the pond that the farmer by the rye had efficiently drained.
    (b) Daniel admired the pond that the farmer by the rye had efficiently harvested.
    (c) Daniel admired the pond that the farmer by the garage had efficiently drained.
    (d) Daniel admired the pond that the farmer by the garage had efficiently harvested.

40. (a) Helen remembered the present that the couple by the willow had carefully picked.
    (b) Helen remembered the present that the couple by the willow had carefully planted.
    (c) Helen remembered the present that the couple by the river had carefully picked.
    (d) Helen remembered the present that the couple by the river had carefully planted.

5.3.7 Experiment 11 stimuli

Below are the experimental items from Experiment 11. Each item consists of the plausible noun/plausible distractor condition in (a), the plausible noun/implausible distractor condition in (b), the implausible noun/plausible distractor condition in (c), the implausible noun/implausible distractor condition in (d), and the generated foil in (e).

1. (a) The dog that chased the coyote was furiously barking all night.
   (b) The dog that chased the ball was furiously barking all night.
   (c) The duck that chased the coyote was furiously barking all night.
   (d) The duck that chased the ball was furiously barking all night.
   (e) x-x-x rely anti heresy gift adhere pre curricula empower join sorry.
2. (a) The wolves that trailed the coyotes were loudly howling last night.
   (b) The wolves that trailed the rabbits were loudly howling last night.
   (c) The hunters that trailed the coyotes were loudly howling last night.
   (d) The hunters that trailed the rabbits were loudly howling last night.
   (e) x-x-x behave jobs alumina guy chaired yeah thesis artisan cent thank.
3. (a) The lamb that followed the goat was continuously bleating yesterday afternoon.
   (b) The lamb that followed the car was continuously bleating yesterday afternoon.
   (c) The dog that followed the goat was continuously bleating yesterday afternoon.
   (d) The dog that followed the car was continuously bleating yesterday afternoon.
   (e) x-x-x spam glad commerce miss ally pre acquisitions imprints guarantee framework.
4. (a) The cow that feared the bull was incessantly mooing all morning.
   (b) The cow that feared the goose was incessantly mooing all morning.
   (c) The girl that feared the bull was incessantly mooing all morning.
   (d) The girl that feared the goose was incessantly mooing all morning.
   (e) x-x-x else cool citing note sided page subtraction regnal pre website.
5. (a) The flies that attacked the bees were annoyingly buzzing all day.

(b) The flies that attacked the rubbish were annoyingly buzzing all day.

(c) The badgers that attacked the bees were annoyingly buzzing all day.

(d) The badgers that attacked the rubbish were annoyingly buzzing all day.

(e) x-x-x heaps yeah treasury glad refund okay orchestras abolish buy sir.

6. (a) The cat that ate the snake was scarily hissing at the man with the hat.

(b) The cat that ate the frog was scarily hissing at the man with the hat.

(c) The crow that ate the snake was scarily hissing at the man with the hat.

(d) The crow that ate the frog was scarily hissing at the man with the hat.

(e) x-x-x nets anti troy fund cyber hate cargoes pertain sale app won cash meet tier.

7. (a) The author that disliked the journalist had carefully written a short novel about nature.

(b) The author that disliked the rain had carefully written a short novel about nature.

(c) The cat that disliked the journalist had carefully written a short novel about nature.

(d) The cat that disliked the rain had carefully written a short novel about nature.

(e) x-x-x anyway anti parkland glad whereas non departure disease miss shall fired china senate.

8. (a) The tailor that hired the seamstress had deftly sewn a silk dress for the client.

(b) The tailor that hired the car had deftly sewn a silk dress for the client.

(c) The company that hired the seamstress had deftly sewn a silk dress for the client.

(d) The company that hired the car had deftly sewn a silk dress for the client.

(e) x-x-x afford okay rocky miss mph pre signer loch oh sued folks sad seem depend.

9. (a) The porter that hired the movers had patiently lifted the heavy suitcase with the equipment.

(b) The porter that hired the taxi had patiently lifted the heavy suitcase with the equipment.

(c) The business that hired the movers had patiently lifted the heavy suitcase with the equipment.

(d) The business that hired the taxi had patiently lifted the heavy suitcase with the equipment.

(e) x-x-x whatever yeah sandy hear hinted pre adversary novels miss teach timeless glad ago regarding.

10. (a) The artist that inspired the painter had always coloured bright portraits in oil.

(b) The artist that inspired the exhibition had always coloured bright portraits in oil.

(c) The scenery that inspired the painter had always coloured bright portraits in oil.

(d) The scenery that inspired the exhibition had always coloured bright portraits in oil.

(e) x-x-x topped hate democrat app shipped dad senate altitude commit nonprofit am lose.

11. (a) The student that specialized in the painter had cautiously sharpened the drawing pencil earlier.

(b) The student that specialized in the rennaissance had cautiously sharpened the drawing pencil earlier.

(c) The museum that specialized in the painter had cautiously sharpened the drawing pencil earlier.

(d) The museum that specialized in the rennaissance had cautiously sharpened the drawing pencil earlier.

(e) x-x-x anyway nice arbitration buy yeah annulled glad campground kilometre kill replace anthem nations.

12. (a) The nurse that assisted the doctor had immediately injected the patient with insulin yesterday.

(b) The nurse that assisted the investigation had immediately injected the patient with insulin yesterday.

(c) The charity that assisted the doctor had immediately injected the patient with insulin yesterday.

(d) The charity that assisted the investigation had immediately injected the patient with insulin yesterday.

(e) x-x-x ignored cool infinity cent thinks pre coincidence beverage glad depends hate plateau celebrate.

13. (a) The assassin that followed the terrorists had already kidnapped the next victim by the evening.

(b) The assassin that followed the parrot had already kidnapped the next victim by the evening.

(c) The dog that followed the terrorists had already kidnapped the next victim by the evening.

(d) The dog that followed the parrot had already kidnapped the next victim by the evening.

(e) x-x-x sided glad sciences hear dismiss miss edition exemption cent lose define ways seem compete.

14. (a) The judge that quarrelled with the magistrate had sentenced the convict to six months.

(b) The judge that quarrelled with the neighbours had sentenced the convict to six months.

(c) The electrician that quarrelled with the magistrate had sentenced the convict to six months.

(d) The electrician that quarrelled with the neighbours had sentenced the convict to six months.

(e) x-x-x tightened holy carcinogen tour push recommends mom algorithm pick denotes okay app nobody.

15. (a) The dressmaker that collaborated with the embroiderer had meticulously woven a silk scarf for her daughter.

(b) The dressmaker that collaborated with the theatre had meticulously woven a silk scarf for her daughter.

(c) The company that collaborated with the embroiderer had meticulously woven a silk scarf for her daughter.

(d) The company that collaborated with the theatre had meticulously woven a silk scarf for her daughter.

(e) x-x-x exporters glad manslaughter hear miss aversion nice principality relic sell outs bytes vote yes explains.

16. (a) The journalist that befriended the writer had reluctantly edited his report last night.

(b) The journalist that befriended the shepherd had reluctantly edited his report last night.

(c) The dancer that befriended the writer had reluctantly edited his report last night.

(d) The dancer that befriended the shepherd had reluctantly edited his report last night.

(e) x-x-x promoting anti divisional else ignore okay liquidation metric lose happen soul worry.

17. (a) The gardener that admired the horticulturist had previously trellised the rose vines in his garden.

(b) The gardener that admired the view had previously trellised the rose vines in his garden.

(c) The guitarist that admired the horticulturist had previously trellised the rose vines in his garden.

(d) The guitarist that admired the view had previously trellised the rose vines in his garden.

(e) x-x-x authored holy mergers app capitulated sir percentage diacritic okay fail jails yeah lose occurs.

18. (a) The physician that consulted the surgeon had expertly cauterised the wart earlier that morning.

(b) The physician that consulted the advisor had expertly cauterised the wart earlier that morning.

(c) The businessman that consulted the surgeon had expertly cauterised the wart earlier that morning.

(d) The businessman that consulted the advisor had expertly cauterised the wart earlier that morning.

(e) x-x-x facilitate nice subtitles jack enhance fun turnpike polyhedron mid vied awesome miss anymore.

19. (a) The goldsmith that yelled at the metalworker had just welded an old ring for his friend.

(b) The goldsmith that yelled at the crowd had just welded an old ring for his friend.

(c) The passenger that yelled at the metalworker had just welded an old ring for his friend.

(d) The passenger that yelled at the crowd had just welded an old ring for his friend.

(e) x-x-x shortages nice nausea lady grow homeostasis yeah jack botany draw seem info hate met happen.

20. (a) The blacksmith that greeted the welder had skillfully tempered the steel for the sword.

(b) The blacksmith that greeted the choir had skillfully tempered the steel for the sword.

(c) The chef that greeted the welder had skillfully tempered the steel for the sword.

(d) The chef that greeted the choir had skillfully tempered the steel for the sword.

(e) x-x-x mediated hate seismic seat devoid pre invariance vanguard miss filed sir than sends.

21. (a) The policeman that despised the sheriff had finally arrested the famous thief yesteray.

(b) The policeman that despised the band had finally arrested the famous thief yesteray.

(c) The baker that despised the sheriff had finally arrested the famous thief yesteray.

(d) The baker that despised the band had finally arrested the famous thief yesteray.

(e) x-x-x secondly holy allusion miss connect mid android expenses push speaks lacks isometry.

22. (a) The army that supported the regime had lately incarcerated many protesters.

(b) The army that supported the parade had lately incarcerated many protesters.

(c) The fundraiser that supported the regime had lately incarcerated many protesters.

(d) The fundraiser that supported the parade had lately incarcerated many protesters.

(e) x-x-x tolerate okay screaming guy meters pre matrix exaggeration wish disclosure.

23. (a) The spy that followed the agent had secretly poisoned the stubborn minister during a meal.

(b) The spy that followed the car had secretly poisoned the stubborn minister during a meal.

(c) The sheep that followed the agent had secretly poisoned the stubborn minister during a meal.

(d) The sheep that followed the car had secretly poisoned the stubborn minister during a meal.

(e) x-x-x spam luck sciences guy hear cent brackets invaders wish ratified consider happen guys rely.

24. (a) The officer that liked the reporter had immediately interviewed the witnesses at the scene.

(b) The officer that liked the walk had immediately interviewed the witnesses at the scene.

(c) The dog that liked the reporter had immediately interviewed the witnesses at the scene.

(d) The dog that liked the walk had immediately interviewed the witnesses at the scene.

(e) x-x-x apart anti roman hour denied pre corporation enterprises miss subscribe sea cent opens.

25. (a) The radiologist that waved to the dentist had previously x-rayed two patients at the clinic.

(b) The radiologist that waved to the singer had previously x-rayed two patients at the clinic.

(c) The plumber that waved to the dentist had previously x-rayed two patients at the clinic.

(d) The plumber that waved to the singer had previously x-rayed two patients at the clinic.

(e) x-x-x scavengers glad undue pre anti vitamin laws inequality bauxite app republic fund dad capita.

26. (a) The carpenter that hired the woodworker had carefully sanded the table last week.

(b) The carpenter that hired the car had carefully sanded the table last week.

(c) The choreographer that hired the woodworker had carefully sanded the table last week.

(d) The choreographer that hired the car had carefully sanded the table last week.

(e) x-x-x irrespective holy fever trip hemorrhage mid paragraph grotto sir shall hate yeah.

27. (a) The nurse that liked the surgeon had carefully sutured the incision from the cut.

(b) The nurse that liked the umbrella had carefully sutured the incision from the cut.

(c) The child that liked the surgeon had carefully sutured the incision from the cut.

(d) The child that liked the umbrella had carefully sutured the incision from the cut.

(e) x-x-x yours cent weird app conclude non emissions sideman pre brokered hate fund nor.

28. (a) The decorator that argued with the workman had already plastered the walls of the office.

(b) The decorator that argued with the fisherman had already plastered the walls of the office.

(c) The chef that argued with the workman had already plastered the walls of the office.

(d) The chef that argued with the fisherman had already plastered the walls of the office.

(e) x-x-x oversee anti tenure cent yeah waivers mid anymore candidacy yes ought pre sir happen.

29. (a) The spy that tracked the agent had succesfully assassinated the foreign president before the meal.

(b) The spy that tracked the teacher had succesfully assassinated the foreign president before the meal.

(c) The musician that tracked the agent had succesfully assassinated the foreign president before the meal.

(d) The musician that tracked the teacher had succesfully assassinated the foreign president before the meal.

(e) x-x-x spam feet spoiler wish aware pre merchantman invitational miss suppose certainly happen cat rely.

30. (a) The craftsman that gazed at the smith had wearily forged a horseshoe that morning.

    (b) The craftsman that gazed at the pigeon had wearily forged a horseshoe that morning.

    (c) The baby that gazed at the smith had wearily forged a horseshoe that morning.

    (d) The baby that gazed at the pigeon had wearily forged a horseshoe that morning.

    (e) x-x-x consumed luck conic jack anti align sir dukedom hacker miss convening glad website.

31. (a) The ballerina that dated the dancer had gracefully pirouetted during the performance.

    (b) The ballerina that dated the wrestler had gracefully pirouetted during the performance.

    (c) The builder that dated the dancer had gracefully pirouetted during the performance.

    (d) The builder that dated the wrestler had gracefully pirouetted during the performance.

    (e) x-x-x plagued anti vegan hate conceded fun capacitors metastable senate hear subscribers.

32. (a) The runner that travelled with the swimmers had passionately competed in the triathlon.

    (b) The runner that travelled with the suitcases had passionately competed in the triathlon.

    (c) The parrot that travelled with the swimmers had passionately competed in the triathlon.

    (d) The parrot that travelled with the suitcases had passionately competed in the triathlon.

    (e) x-x-x expire jobs fisheries soul app amounting mid governorship nutrient lord glad superiors.

33. (a) The thief that led the gang had recently robbed the bank next to the school.

    (b) The thief that led the parade had recently robbed the bank next to the school.

    (c) The sign that led the gang had recently robbed the bank next to the school.

(d) The sign that led the parade had recently robbed the bank next to the school.

(e) x-x-x sided luck cent wish ours yeah download metric met seem mean yes kid thinks.

34. (a) The kidnapper that liked the pirates had eventually ransomed the wealthy investor yesterday night.

(b) The kidnapper that liked the pool had eventually ransomed the wealthy investor yesterday night.

(c) The toddler that liked the pirates had eventually ransomed the wealthy investor yesterday night.

(d) The toddler that liked the pool had eventually ransomed the wealthy investor yesterday night.

(e) x-x-x collided nice stats sir cancel anti researcher peerages glad tweeted analyzed recommend apply.

35. (a) The agent that chased the gangster had finally extorted a long confession from the informant.

(b) The agent that chased the birds had finally extorted a long confession from the informant.

(c) The dog that chased the gangster had finally extorted a long confession from the informant.

(d) The dog that chased the birds had finally extorted a long confession from the informant.

(e) x-x-x apart anti lagoon wish regret pre discuss headland app yeah incentives cool push coincides.

36. (a) The criminologist that disliked the detective had wisely conducted an investigation into the accident.

(b) The criminologist that disliked the noise had wisely conducted an investigation into the accident.

(c) The cat that disliked the detective had wisely conducted an investigation into the accident.

(d) The cat that disliked the noise had wisely conducted an investigation into the accident.

(e) x-x-x deflections glad delaying else submit holy bidder southeast app personalities rose cent maintain.

5.3.8 Experiment 12 stimuli

Below are the experimental items from Experiment 12. Each item consists of the plausible noun/plausible distractor condition in (a), the plausible noun/implausible distractor condition in (b), the implausible noun/plausible distractor condition in (c), and the implausible noun/implausible distractor condition in (d).

1. (a) The dog that chased the coyote was furiously barking all night yesterday.
   (b) The dog that chased the ball was furiously barking all night yesterday.
   (c) The duck that chased the coyote was furiously barking all night yesterday.
   (d) The duck that chased the ball was furiously barking all night yesterday.

2. (a) The wolves that trailed the coyotes were loudly howling late last night.
   (b) The wolves that trailed the rabbits were loudly howling late last night.
   (c) The hunters that trailed the coyotes were loudly howling late last night.
   (d) The hunters that trailed the rabbits were loudly howling late last night.

3. (a) The lamb that followed the goat was continuously bleating in the afternoon.
   (b) The lamb that followed the car was continuously bleating in the afternoon.
   (c) The dog that followed the goat was continuously bleating in the afternoon.
   (d) The dog that followed the car was continuously bleating in the afternoon.

4. (a) The cow that feared the bull was incessantly mooing the whole morning.
   (b) The cow that feared the goose was incessantly mooing the whole morning.
   (c) The girl that feared the bull was incessantly mooing the whole morning.
   (d) The girl that feared the goose was incessantly mooing the whole morning.

5. (a) The flies that attacked the bees were annoyingly buzzing all day today.
   (b) The flies that attacked the rubbish were annoyingly buzzing all day today.
   (c) The badgers that attacked the bees were annoyingly buzzing all day today.
   (d) The badgers that attacked the rubbish were annoyingly buzzing all day today.

6. (a) The cat that ate the snake was scarily hissing at the man.
   (b) The cat that ate the frog was scarily hissing at the man.
   (c) The crow that ate the snake was scarily hissing at the man.

(d) The crow that ate the frog was scarily hissing at the man.

7. (a) The author that disliked the journalist had carefully written a short novel.
   (b) The author that disliked the rain had carefully written a short novel.
   (c) The cat that disliked the journalist had carefully written a short novel.
   (d) The cat that disliked the rain had carefully written a short novel.

8. (a) The tailor that hired the seamstress had deftly sewn a silk dress.
   (b) The tailor that hired the car had deftly sewn a silk dress.
   (c) The company that hired the seamstress had deftly sewn a silk dress.
   (d) The company that hired the car had deftly sewn a silk dress.

9. (a) The porter that hired the movers had patiently lifted the heavy suitcase.
   (b) The porter that hired the taxi had patiently lifted the heavy suitcase.
   (c) The business that hired the movers had patiently lifted the heavy suitcase.
   (d) The business that hired the taxi had patiently lifted the heavy suitcase.

10. (a) The artist that inspired the painter had always coloured bright portraits in oil.
    (b) The artist that inspired the exhibition had always coloured bright portraits in oil.
    (c) The scenery that inspired the painter had always coloured bright portraits in oil.
    (d) The scenery that inspired the exhibition had always coloured bright portraits in oil.

11. (a) The student that specialised in the painter had cautiously sharpened the pencil.
    (b) The student that specialised in the rennaissance had cautiously sharpened the pencil.
    (c) The museum that specialised in the painter had cautiously sharpened the pencil.
    (d) The museum that specialised in the rennaissance had cautiously sharpened the pencil.

12. (a) The nurse that assisted the doctor had immediately injected the patient with insulin.

(b) The nurse that assisted the investigation had immediately injected the patient with insulin.

(c) The charity that assisted the doctor had immediately injected the patient with insulin.

(d) The charity that assisted the investigation had immediately injected the patient with insulin.

13. (a) The assassin that followed the terrorists had already kidnapped the next victim.

(b) The assassin that followed the parrot had already kidnapped the next victim.

(c) The dog that followed the terrorists had already kidnapped the next victim.

(d) The dog that followed the parrot had already kidnapped the next victim.

14. (a) The judge that quarrelled with the magistrate had sentenced the convict to six months.

(b) The judge that quarrelled with the neighbours had sentenced the convict to six months.

(c) The electrician that quarrelled with the magistrate had sentenced the convict to six months.

(d) The electrician that quarrelled with the neighbours had sentenced the convict to six months.

15. (a) The dressmaker that collaborated with the embroiderer had meticulously woven a silk scarf.

(b) The dressmaker that collaborated with the theatre had meticulously woven a silk scarf.

(c) The company that collaborated with the embroiderer had meticulously woven a silk scarf.

(d) The company that collaborated with the theatre had meticulously woven a silk scarf.

16. (a) The journalist that befriended the write had reluctantly edited his report last night.

(b) The journalist that befriended the shepherd had reluctantly edited his report last night.

(c) The dancer that befriended the write had reluctantly edited his report last night.

(d) The dancer that befriended the shepherd had reluctantly edited his report last night.

17. (a) The gardener that admired the horticulturist had previously trellised the rose vines.

(b) The gardener that admired the view had previously trellised the rose vines.

(c) The guitarist that admired the horticulturist had previously trellised the rose vines.

(d) The guitarist that admired the view had previously trellised the rose vines.

18. (a) The physician that consulted the surgeon had expertly cauterised the wart earlier.

(b) The physician that consulted the advisor had expertly cauterised the wart earlier.

(c) The businessman that consulted the surgeon had expertly cauterised the wart earlier.

(d) The businessman that consulted the advisor had expertly cauterised the wart earlier.

19. (a) The goldsmith yelled at the metalworker had just welded an old ring.

(b) The goldsmith yelled at the crowd had just welded an old ring.

(c) The passenger yelled at the metalworker had just welded an old ring.

(d) The passenger yelled at the crowd had just welded an old ring.

20. (a) The blacksmith that greeted the welder had skillfully tempered the steel for the sword.

(b) The blacksmith that greeted the choir had skillfully tempered the steel for the sword.

(c) The chef that greeted the welder had skillfully tempered the steel for the sword.

(d) The chef that greeted the choir had skillfully tempered the steel for the sword.

21. (a) The policeman that despised the sheriff had finally arrested the famous thief.

(b) The policeman that despised the band had finally arrested the famous thief.

(c) The baker that despised the sheriff had finally arrested the famous thief.

(d) The baker that despised the band had finally arrested the famous thief.

22. (a) The army that supported the regime had lately incarcerated many young protesters.

(b) The army that supported the parade had lately incarcerated many young protesters.

(c) The fundraiser that supported the regime had lately incarcerated many young protesters.

(d) The fundraiser that supported the parade had lately incarcerated many young protesters.

23. (a) The spy that followed the agent had secretly poisoned the stubborn minister.

(b) The spy that followed the car had secretly poisoned the stubborn minister.

(c) The sheep that followed the agent had secretly poisoned the stubborn minister.

(d) The sheep that followed the car had secretly poisoned the stubborn minister.

24. (a) The officer that liked the reporter had immediately interviewed the terrified witnesses.

(b) The officer that liked the walk had immediately interviewed the terrified witnesses.

(c) The dog that liked the reporter had immediately interviewed the terrified witnesses.

(d) The dog that liked the walk had immediately interviewed the terrified witnesses.

25. (a) The radiologist that waved to the dentist had previously x-rayed two patients.

(b) The radiologist that waved to the singer had previously x-rayed two patients.

(c) The plumber that waved to the dentist had previously x-rayed two patients.

(d) The plumber that waved to the singer had previously x-rayed two patients.

26. (a) The carpenter that hired the woodworker had carefully sanded the table last week.

(b) The carpenter that hired the car had carefully sanded the table last week.

(c) The choreographer that hired the woodworker had carefully sanded the table last week.

(d) The choreographer that hired the car had carefully sanded the table last week.

27. (a) The nurse that liked the surgeon had carefully sutured the incision from the cut.

(b) The nurse that liked the umbrella had carefully sutured the incision from the cut.

(c) The child that liked the surgeon had carefully sutured the incision from the cut.

(d) The child that liked the umbrella had carefully sutured the incision from the cut.

28. (a) The decorator that argued with the workman had already plastered the walls.

(b) The decorator that argued with the fisherman had already plastered the walls.

(c) The chef that argued with the workman had already plastered the walls.

(d) The chef that argued with the fisherman had already plastered the walls.

29. (a) The spy that tracked the agent had succesfully assassinated the foreign president.

(b) The spy that tracked the teacher had succesfully assassinated the foreign president.

(c) The musician that tracked the agent had succesfully assassinated the foreign president.

(d) The musician that tracked the teacher had succesfully assassinated the foreign president.

30. (a) The craftsman that gazed at the smith had wearily forged a horseshoe.

(b) The craftsman that gazed at the pigeon had wearily forged a horseshoe.

(c) The baby that gazed at the smith had wearily forged a horseshoe.

(d) The baby that gazed at the pigeon had wearily forged a horseshoe.

31. (a) The ballerina that dated the dancer had gracefully pirouetted during the performance.

(b) The ballerina that dated the wrestler had gracefully pirouetted during the performance.

(c) The builder that dated the dancer had gracefully pirouetted during the performance.

(d) The builder that dated the wrestler had gracefully pirouetted during the performance.

32. (a) The runner that travelled with the swimmers had passionately competed in the triathlon.

(b) The runner that travelled with the suitcases had passionately competed in the triathlon.

(c) The parrot that travelled with the swimmers had passionately competed in the triathlon.

(d) The parrot that travelled with the suitcases had passionately competed in the triathlon.

33. (a) The thief that led the gang had recently robbed the bank in town.

(b) The thief that led the parade had recently robbed the bank in town.

(c) The sign that led the gang had recently robbed the bank in town.

(d) The sign that led the parade had recently robbed the bank in town.

34. (a) The kidnapper that liked the pirates had eventually ransomed the wealthy investor.

(b) The kidnapper that liked the pool had eventually ransomed the wealthy investor.

(c) The toddler that liked the pirates had eventually ransomed the wealthy investor.

(d) The toddler that liked the pool had eventually ransomed the wealthy investor.

35. (a) The agent that chased the gangster had finally extorted a long confession.

(b) The agent that chased the birds had finally extorted a long confession.

(c) The dog that chased the gangster had finally extorted a long confession.

(d) The dog that chased the birds had finally extorted a long confession.

36. (a) The criminologist that disliked the detective had wisely conducted an investigation into the accident.

(b) The criminologist that disliked the noise had wisely conducted an investigation into the accident.

(c) The cat that disliked the detective had wisely conducted an investigation into the accident.

(d) The cat that disliked the noise had wisely conducted an investigation into the accident.

Bibliography

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Anderson, J. R. (2013a). *The architecture of cognition*. Psychology Press.

Anderson, J. R. (2013b). *Language, memory, and thought*. Psychology Press.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*(4), 341–380.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive psychology*, *30*(3), 221–256.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods*, *52*(1), 388–407.

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of memory and language*, *51*(1), 23–54.

Ariel, M. (1990). Accessing noun-phrase antecedents. vol. 96. londres: Routlege. *Linguistics*, 113–118.

Atkinson, E., Wagers, M. W., Lidz, J., Phillips, C., & Omaki, A. (2018). Developing incrementality in filler-gap dependency processing. *Cognition*, *179*, 132–149.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (pp. 89–195, Vol. 2). Elsevier.

Audley, R., & Pike, A. (1965). Some alternative stochastic models of choice 1. *British Journal of Mathematical and Statistical Psychology*, *18*(2), 207–225.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.

Baddeley, A. (2002). Fractionating the central executive. *Principles of frontal lobe function*, 246–260.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature reviews neuroscience*, *4*(10), 829–839.

Baddeley, A., Chincotta, D., Stafford, L., & Turk, D. (2002). Is the word length effect in stm entirely attributable to output delay? evidence from serial recognition. *The Quarterly Journal of Experimental Psychology Section A*, *55*(2), 353–369.

Baddeley, A., Logie, R., Bressi, S., Sala, S. D., & Spinnler, H. (1986). Dementia and working memory. *The Quarterly Journal of Experimental Psychology Section A*, *38*(4), 603–618.

Baddeley, A., & Wilson, B. (1985). Phonological coding and short-term memory in patients without speech. *Journal of memory and language*, *24*(4), 490–502.

Baddeley, A. D. (1966a). The capacity for generating information by randomization. *The Quarterly journal of experimental psychology*, *18*(2), 119–129.

Baddeley, A. D. (1966b). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly journal of experimental psychology*, *18*(4), 362–365.

Baddeley, A. D., BRESSI, S., Della Sala, S., Logie, R., & SPINNLER, H. (1991). The decline of working memory in alzheimer's disease: A longitudinal study. *Brain*, *114*(6), 2521–2542.

Baddeley, A. D., & Graham, J. (1974). Hitch (1974). working memory. *The psychology of learning and motivation*, *8*.

Baddeley, A. D., & Lieberman, K. (2017). Spatial working memory. In *Exploring working memory* (pp. 206–223). Routledge.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior*, *14*(6), 575–589.

Baddeley, A. D., & Warrington, E. K. (1970). Amnesia and the distinction between long-and short-term memory. *Journal of verbal learning and verbal behavior*, *9*(2), 176–189.

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in slovak. *Journal of memory and language*, *56*(1), 65–85.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748.

Balogh, J., Zurif, E., Prather, P., Swinney, D., & Finkel, L. (1998). Gap-filling and end-of-sentence effects in real-time language processing: Implications for modeling sentence comprehension in aphasia. *Brain and Language*, *61*(2), 169–182.

Basso, A., Spinnler, H., Vallar, G., & Zanobio, M. E. (1982). Left hemisphere damage and selective impairment of auditory verbal short-term memory. a case study. *Neuropsychologia*, *20*(3), 263–274.

Bishop, D., & Robson, J. (1989). Accurate non-word spelling despite congenital inability to speak: Phoneme—grapheme conversion does not require subvocal articulation. *British Journal of Psychology*, *80*(1), 1–13.

Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in english number agreement. *Language and Cognitive Processes*, *8*(1), 57–99.

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive psychology*, *43*(2), 83–128.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive psychology*, *23*(1), 45–93.

Bock, K., Nicol, J., & Cutting, J. C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and language*, *40*(3), 330–346.

Boyce, V., Futrell, R., & Levy, R. P. (2020). Maze made easy: Better and easier measurement of incremental processing difficulty. *Journal of Memory and Language*, *111*, 104082.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. (1992). Manipulation of visual mental images in children and adults. *Journal of Experimental Child Psychology*, *53*(3), 300–312.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly journal of experimental psychology*, *10*(1), 12–21.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive psychology*, *57*(3), 153–178.

Builder, S. R. E. (2007). Computer software.

Bürkner, P.-C. (2017a). Advanced bayesian multilevel modeling with the r package brms. *arXiv preprint arXiv:1705.11123*.

Bürkner, P.-C. (2017b). Brms: An r package for bayesian multilevel models using stan. *Journal of statistical software*, *80*, 1–28.

Caplan, D., & Waters, G. (2013). Memory mechanisms supporting syntactic comprehension. *Psychonomic bulletin & review*, *20*, 243–268.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and brain Sciences*, *22*(1), 77–94.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76*(1).

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the raven progressive matrices test. *Psychological review*, *97*(3), 404.

Chomsky, N. (1981). Lectures on government and binding theory. *Dordrecht: Foris*.

Chow, W.-Y., Smith, C., Lau, E., & Phillips, C. (2016). A "bag-of-arguments" mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, *31*(5), 577–596.

Clifton Jr, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In *Linguistic structure in language processing* (pp. 273–317). Springer.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British journal of Psychology*, *55*(1), 75–84.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British journal of psychology*, *55*(4), 429–432.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, *24*(1), 87–114.

Cowan, N. (2012). *Working memory capacity*. Psychology press.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of memory and language*, *31*(1), 1–17.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers. *Natural language parsing: Psychological, computational, and theoretical perspectives*, 94–128.

Cunnings, I., & Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory and Language*, *75*, 117–139.

Cunnings, I., & Sturt, P. (2018). Retrieval interference and semantic interpretation. *Journal of Memory and Language*, *102*, 16–27.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic bulletin & review*, *3*(4), 422–433.

De Renzi, E., & Nichelli, P. (1975). Verbal and non-verbal short-term memory impairment following hemispheric damage. *Cortex*, *11*(4), 341–354.

De Vincenzi, M. (1991). "filler-gap dependencies in a null subject language: Referential and nonreferential whs.". *Journal of psycholinguistic research*, *20*, 197–213.

De Vincenzi, M. (2000). "cross-linguistic psycholinguistics". *Architectures and mechanisms for language processing*, 282–300.

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*(2), 85–103.

Drummond, A. (2013). Ibex farm. *Online server: http://spellout. net/ibexfarm*, *831*.

Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *Journal of Memory and language*, *36*(2), 147–164.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 972.

Farah, M. J., Hammond, K. M., Levine, D. N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive psychology*, *20*(4), 439–462.

Farmer, E. W., Berman, J. V., & Fletcher, Y. L. (1986). Evidence for a visuo-spatial scratch-pad in working memory. *The Quarterly Journal of Experimental Psychology Section A*, *38*(4), 675–688.

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.

Fellbaum, C. (2010). Wordnet. In *Theory and applications of ontology: Computer applications* (pp. 231–243). Springer.

Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of memory and language*, *25*(3), 348–368.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic inquiry*, *9*(3), 427–473.

Fodor, J. D., Ni, W., Crain, S., & Shankweiler, D. (1996). Tasks and timing in the perception of linguistic anomaly. *Journal of Psycholinguistic Research*, *25*(1), 25–57.

Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior research methods*, *41*(1), 163–171.

Franck, J., Lassi, G., Frauenfelder, U. H., & Rizzi, L. (2006). Agreement and movement: A syntactic analysis of attraction. *Cognition*, *101*(1), 173–216.

Franck, J., Soare, G., Frauenfelder, U. H., & Rizzi, L. (2010). Object interference in subject–verb agreement: The role of intermediate traces of movement. *Journal of memory and language*, *62*(2), 166–182.

Franck, J., Vigliocco, G., Antón-Méndez, I., Collina, S., & Frauenfelder, U. H. (2008). The interplay of syntax and form in sentence production: A cross-linguistic study of form effects on agreement. *Language and Cognitive Processes*, *23*(3), 329–374.

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in french and english: The role of syntactic hierarchy. *Language and cognitive processes*, *17*(4), 371–404.

Frazier, L. (1987a). Sentence processing: A tutorial review. *Attention and performance XII*, 559–586.

Frazier, L. (1987b). Syntactic processing: Evidence from dutch. *Natural Language & Linguistic Theory*, *5*(4), 519–559.

Frazier, L., & Clifton Jr, C. (1989). Successive cyclicity in the grammar and the parser. *Language and cognitive processes*, *4*(2), 93–126.

Frazier, L., & d'Arcais, G. B. F. (1989). Filler driven parsing: A study of gap filling in dutch. *Journal of memory and language*, *28*(3), 331–344.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325.

Fujita, H., & Cunnings, I. (2022). Interference and filler-gap dependency formation in native and non-native language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(5), 702.

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive science*, *44*(3), e12814.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of memory and language*, *37*(1), 58–93.

Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of psycholinguistic research*, *18*, 51–60.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gibson, E., et al. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, *2000*, 95–126.

Gibson, E., & Ko, K. (1998). An integration-based theory of computational resources in sentence comprehension. *Fourth Architectures and Mechanisms in Language Processing Conference, University of Freiburg, Germany*.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of experimental psychology: learning, memory, and cognition*, *27*(6), 1411.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of memory and Language*, *51*(1), 97–114.

Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of experimental psychology: Learning, Memory, and Cognition*, *32*(6), 1304.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological science*, *13*(5), 425–430.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive science*, *29*(2), 261–290.

Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive psychology*, *110*, 70–104.

Hanley, J. R., Young, A. W., & Pearson, N. A. (1991). Impairment of the visuo-spatial sketch pad. *The quarterly Journal of experimental Psychology*, *43*(1), 101–125.

Hartsuiker, R. J., Kolk, H. H., & Huinck, W. J. (1999). Agrammatic production of subject–verb agreement: The effect of conceptual number. *Brain and language*, *69*(2), 119–160.

Hartsuiker, R. J., Schriefers, H. J., Bock, K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject-verb agreement. *Memory & cognition*, *31*(8), 1316–1326.

Häussler, J. (2009). The emergence of attraction errors during sentence comprehension.

Hebb, D. (1949). Organization of behavior. new york: Wiley. *J. Clin. Psychol*, *6*(3), 335–307.

Hill, R. L., & Murray, W. S. (2000). Commas and spaces: Effects of punctuation on eye movements and sentence parsing. In *Reading as a perceptual process* (pp. 565–589). Elsevier.

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in chinese. *Cognition*, *90*(1), 3–27.

Jäger, L. A., Benz, L., Roeser, J., Dillon, B. W., & Vasishth, S. (2015). Teasing apart retrieval and encoding interference in the processing of anaphors. *Frontiers in psychology*, *6*, 506.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339.

Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annu. Rev. Psychol.*, *59*, 193–224.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, *87*(4), 329.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological review*, *99*(1), 122.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of experimental psychology: General*, *111*(2), 228.

Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject-verb dependencies: An erp study. *Journal of Psycholinguistic Research*, *31*(2), 165–193.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and language*, *49*(1), 133–156.

Kemper, S., Crow, A., & Kemtes, K. (2004). Eye-fixation patterns of high-and low-span young and older adults: Down the garden path and back again. *Psychology and Aging*, *19*(1), 157.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods*, *42*(3), 627–633.

Klauer, K. C., & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of experimental psychology: General*, *133*(3), 355.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain research*, *1146*, 23–49.

Kush, D., & Phillips, C. (2014). Local anaphor licensing in an sov language: Implications for retrieval strategies. *Frontiers in psychology*, *5*, 1252.

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in spanish comprehension. *Journal of Memory and Language*, *82*, 133–149.

Lange, E. B., & Oberauer, K. (2005). Overwriting of phonemic features in serial recall. *Memory*, *13*(3-4), 333–339.

Laurinavichyute, A., Jäger, L. A., Akinina, Y., Roß, J., & Dragoy, O. (2017). Retrieval and encoding interference: Cross-linguistic evidence from anaphor processing. *Frontiers in psychology*, *8*, 965.

Laurinavichyute, A., & Von der Malsburg, T. (2024). Agreement attraction in grammatical sentences and the role of the task. *Journal of Memory and Language*, *137*, 104525.

Laurinavichyute, A., & von der Malsburg, T. (2022). Semantic attraction in sentence comprehension. *Cognitive Science*, *46*(2), e13086.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of psycholinguistic research*, *25*(1), 93–115.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science*, *29*(3), 375–419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in cognitive sciences*, *10*(10), 447–454.

Liao, C.-H., Lau, E., & Chow, W.-Y. (2022). Towards a processing model for argument-verb computations in online sentence comprehension. *Journal of Memory and Language*, *126*, 104350.

Logačev, P., & Vasishth, S. (2016). A multiple-channel model of task-dependent ambiguity resolution in sentence comprehension. *Cognitive Science*, *40*(2), 266–298.

Logie, R. H. (1986). Visuo-spatial processing in working memory. *The Quarterly Journal of Experimental Psychology Section A*, *38*(2), 229–247.

Logie, R. H., Cocchini, G., Delia Sala, S., & Baddeley, A. D. (2004). Is there a specific executive capacity for dual task coordination? evidence from alzheimer's disease. *Neuropsychology*, *18*(3), 504.

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta psychologica*, *75*(1), 55–74.

Lorimor, H., Bock, K., Zalkind, E., Sheyman, A., & Beard, R. (2008). Agreement and attraction in russian. *Language and cognitive processes*, *23*(6), 769–799.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive psychology*, *24*(1), 56–98.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological review*, *101*(4), 676.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*(5417), 522–523.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, *189*(4198), 226–228.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of psycholinguistic research*, *29*, 111–123.

McElree, B. (2006). Accessing recent events. *Psychology of learning and motivation*, *46*, 155–200.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of memory and language*, *48*(1), 67–91.

Mertzen, D., Laurinavichyute, A., Dillon, B., Engbert, R., & Vasishth, S. (2020). A cross-linguistic investigation of proactive, similarity-based retrieval interference in sentence comprehension: No support from english, german and russian eye-tracking data. *German and Russian eye-tracking data*.

Meyer, A. S., & Bock, K. (1999). Representations and processes in the production of pronouns: Some perspectives from dutch. *Journal of Memory and Language*, *41*(2), 281–301.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Miller, G. A. (1995). Wordnet: A lexical database for english. *Communications of the ACM*, *38*(11), 39–41.

Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, *2*(3), 217–228.

Milner, B. A. (1966). Amnesia following operation on the temporal lobes. *Neurocase*, *2*.

Mitchell, D. C., & Green, D. W. (1978). The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology*, *30*(4), 609–636.

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *21*(3), 263.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & cognition*, *18*, 251–269.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic bulletin & review*, *7*, 403–423.

Ness, T., & Meltzer-Asscher, A. (2017). Working memory in the processing of long-distance dependencies: Interference and filler maintenance. *Journal of psycholinguistic research*, *46*, 1353–1365.

Ness, T., & Meltzer-Asscher, A. (2019). When is the verb a potential gap site? the influence of filler maintenance on the active search for a gap. *Language, Cognition and Neuroscience*, *34*(7), 936–948.

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational ideas—part ii. *Language and Linguistics Compass*, *10*(11), 591–613.

Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using bayesian hierarchical modeling. *Journal of Memory and Language*, *99*, 1–34.

Nicenboim, B., Vasishth, S., Gattei, C., Sigman, M., & Kliegl, R. (2015). Working memory differences in long-distance dependency resolution. *Frontiers in Psychology*, *6*, 312.

Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject–verb agreement processes in comprehension. *Journal of Memory and Language*, *36*(4), 569–587.

Nissen, M. J. (1985). Accessing features and objects: Is location special. *Attention and performance XI*, 205–219.

Norman, D. A., & Shallice, T. (1986). Attention to action. In *Consciousness and self-regulation* (pp. 1–18). Springer.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of memory and language*, *55*(4), 601–626.

O'Hagan, A., & Forster, J. J. (2004). *Kendall's advanced theory of statistics, volume 2b: Bayesian inference* (Vol. 2). Arnold.

Omaki, A., Davidson White, I., Goro, T., Lidz, J., & Phillips, C. (2014). No fear of commitment: Children's incremental interpretation in english and japanese wh-questions. *Language Learning and Development*, *10*(3), 206–233.

Omaki, A., Lau, E. F., Davidson White, I., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in psychology*, *6*, 384.

Osterhout, L., & Swinney, D. A. (1993). On the temporal course of gap-filling during comprehension of verbal passives. *Journal of Psycholinguistic Research*, *22*(2), 273–286.

Page, M., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological review*, *105*(4), 761.

Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Quarterly Journal of Experimental Psychology*, *69*(5), 950–971.

Patwardhan, S. (2006). Pederson, t. *Freely Similarity Package. Aailable on: http://wn-similarity. sourceforge. net*.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and language*, *41*(3), 427–456.

Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of memory and language*, *34*(4), 521–542.

Pedersen, T., Patwardhan, S., Michelizzi, J., et al. (2004). Wordnet:: Similarity-measuring the relatedness of concepts. *AAAI*, *4*, 25–29.

Pederson, T., Patwardhan, S., Banerjee, S., & Michelizzi, J. (2008). Text:: Similarity. *Duluth, Minnesota: University of Minnesota*.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). Erp effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*(3), 407–428.

Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and cognitive processes*, *6*(3), 229–259.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 940.

Pickering, M. J., & Traxler, M. J. (2003). Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. *Language and Cognitive Processes*, *18*(4), 469–503.

Potter, M. C. (1984). Rapid serial visual presentation (rsvp): A method for studying language processing. *New methods in reading comprehension research*, *118*, 91–118.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922.

Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology: Section A*, *53*(4), 1061–1080.

Research, S. (2004). Experiment builder.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*(3), 211.

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing–storage relationship in reading span performance. *Journal of memory and Language*, *50*(4), 425–443.

Servan-Schreiber, E. (1991). The competitive chunking theory: Models of perception, learning, and memory. *Dissertation Abstracts International*, *52*(6-B), 3279.

Shallice, T., et al. (2002). Fractionation of the supervisory system. *Principles of frontal lobe function*, 261–277.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *The Quarterly journal of experimental psychology*, *22*(2), 261–273.

Smith, G., & Vasishth, S. (2020). A principled approach to feature selection in models of sentence processing. *Cognitive science*, *44*(12), e12918.

Smyth, M. M., & Pendleton, L. R. (1990). Space and movement in working memory. *The Quarterly Journal of Experimental Psychology Section A*, *42*(2), 291–304.

Smyth, M. M., & Scholey, K. A. (1992). Determining spatial span: The role of movement time and articulation rate. *The Quarterly Journal of Experimental Psychology Section A*, *45*(3), 479–501.

Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working-memory capacity and syntactic island effects. *Language*, 82–123.

Stabler, E. P. (2015). The finite connectivity of linguistic structure. In *Perspectives on sentence processing* (pp. 303–336). Psychology Press.

Stan Development Team. (2020). RStan: The R interface to Stan [R package version 2.26.24]. https://mc-stan.org/

Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 550.

Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of memory and language*, *60*(2), 308–327.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71–86.

Stowe, L. A. (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and cognitive processes*, *1*(3), 227–245.

Swinney, D., & Zurif, E. (1995). Syntactic processing in aphasia. *Brain and Language*, *50*(2), 225–239.

Tanenhaus, M. K., & Donnenwerth-Nolan, S. (1984). Syntactic context and lexical access. *The Quarterly Journal of Experimental Psychology*, *36*(4), 649–661.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of memory and language*, *76*, 195–215.

Taylor, W. L. (1953). "cloze procedure": A new tool for measuring readability. *Journalism quarterly*, *30*(4), 415–433.

Team R Core. (2020). R: A language and environment for statistical computing (version 4.0. 2). *R Foundation for Statistical Computing*.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454–475.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of memory and language*, *33*(3), 285–318.

Tyler, L. K., & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of verbal learning and verbal behavior*, *16*(6), 683–692.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological review*, *108*(3), 550.

Vallar, G., & Baddeley, A. D. (1984a). Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 151–161.

Vallar, G., & Baddeley, A. D. (1984b). Phonological short-term store, phonological processing and sentence comprehension: A neuropsychological case study. *Cognitive neuropsychology*, *1*(2), 121–141.

van Rijn, H., & R Anderson, J. (2003). Modeling lexical decision as ordinary retrieval.

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 407.

Van Dyke, J. A., & Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Language and linguistics compass*, *6*(4), 193–211.

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.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of memory and language*, *55*(2), 157–166.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of memory and language*, *65*(3), 247–263.

Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In *Reading as a perceptual process* (pp. 621–648). Elsevier.

Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, *45*(2), 225–258.

Van Maanen, L., van Rijn, H., & Taatgen, N. (2012). Race/a: An architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive science*, *36*(1), 62–101.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and erp components. *International journal of psychophysiology*, *83*(2), 176–190.

Vasishth, S., Brüssow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*(4), 685–712.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, *13*(1), 37–58.

Vigliocco, G. (1996). One or more labels on the bottles? notional concord in dutch and french. *Language and Cognitive Processes*, *11*(4), 407–442.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject-verb agreement in spanish and english: Differences in the role of conceptual constraints. *Cognition*, *61*(3), 261–298.

Vigliocco, G., Butterworth, B., & Semenza, C. (1995). Constructing subject-verb agreement in speech: The role of semantic and morphological factors. *Journal of Memory and Language*, *34*(2), 186–215.

Vigliocco, G., & Franck, J. (2001). When sex affects syntax: Contextual influences in sentence production. *Journal of Memory and Language*, *45*(3), 368–390.

Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in psychology*, *9*, 2.

Von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1545–1578.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of memory and language*, *61*(2), 206–237.

Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology*, *67*(7), 1274–1304.

Wagers, M. W., Phillips, C., & Wagers, M. (2012). Going the distance: Memory and decision making in active dependency construction. *Quarterly Journal of Experimental Psychology*.

Wanner, E., Shiner, S., & Kaplan, R. (1978). *Garden paths in relative clauses*. Harvard Univ.

Wanner, H., & Maratsos, M. P. (1978). An atn approach to comprehension. In *Linguistic theory and psychological reality*. mit Press.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*(1), 79–112.

Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and e–z reader. *Cognition*, *111*(1), 132–137.

Warrington, E. K., & Weiskrantz, L. (1968). A study of learning and retention in amnesic patients. *Neuropsychologia*, *6*(3), 283–291.

White, S. J., Warren, T., & Reichle, E. D. (2011). Parafoveal preview during reading: Effects of sentence position. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(4), 1221.

Wickelgren, W. A. (1965). Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, *70*(1), 102.

Wickelgren, W. A. (1968). Sparing of short-term memory in an amnesic patient: Implications for strength theory of memory. *Neuropsychologia*, *6*(3), 235–244.

Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: Eye tracking, moving-window, and maze. *Journal of psycholinguistic research*, *41*(2), 105–128.

Yadav, H., Smith, G., Reich, S., & Vasishth, S. (2023). Number feature distortion modulates cue-based retrieval in reading. *Journal of Memory and Language*, *129*, 104400.

Zehr, J., & Schwarz, F. (2018a). Penncontroller for internet based experiments (ibex). *DOI: https://doi. org/10.17605/OSF. IO/MD832*.

Zehr, J., & Schwarz, F. (2018b). Penncontroller for internet based experiments (ibex).