Retrieval interference and semantic interpretation

Retrieval interference and semantic interpretation

Ian Cunnings
University of Reading, UK

&

Patrick Sturt
University of Edinburgh, UK

Address for correspondence:

Dr Ian Cunnings
School of Psychology and Clinical Language Sciences
University of Reading
Reading, UK
RG6 7BE

Email: i.cunnings@reading.ac.uk
Phone: +44 (0)118 378 5538

Abstract

Similarity-based interference has played an important role in motivating cue-based models of memory retrieval during language comprehension. One example of interference comes from *illusions of grammaticality*, where ungrammatical sentences are perceived as grammatical (e.g. 'the key to the cabinets were rusty'). While such effects indicate interference influences perception of sentence grammaticality, less is known about how interference influences the semantic interpretation assigned to a sentence. We report two reading experiments that manipulated sentence plausibility, rather than grammaticality, as a diagnostic of interference. In both experiments, although reading times were longer for implausible sentences, this plausibility effect was reliably attenuated when a distractor item partially matched the cues at retrieval. We interpret these results as being compatible with the predictions of cue-based parsing. The *illusions of plausibility* that we report indicate that similarity-based retrieval interference has a potent influence on the semantic interpretation that is assigned to a sentence during processing.

Keywords: Sentence processing; memory retrieval; interference; linguistic dependencies; eye-movements; reading

Introduction

Similarity-based retrieval interference has played an important role in models of short term memory (Jonides, Lewis, Nee, Lusting, Berman & Moore, 2008). Research on the comprehension of linguistic dependencies also indicates that similarity-based interference is a determinant of memory retrieval during language comprehension (Lewis, Vasishth & Van Dyke, 2006; Van Dyke & Johns, 2012). According to cue-based models of parsing (e.g. Lewis & Vasishth, 2005; Lewis et al., 2000; McElree, 2000, Van Dyke, 2007), interference arises as a result of the match between the set of cues utilised at retrieval and the number of items in memory that match these cues. One piece of evidence for retrieval interference during language comprehension comes from so-called *illusions of grammaticality*, where ungrammatical sentences are perceived as grammatical (Phillips, Wagers & Lau, 2011). Although such illusions suggest interference can influence perception of sentence grammaticality, it is less clear how this type of interference may influence the semantic interpretation that is assigned to a sentence as it unfolds.

The aim of this study was to investigate how retrieval interference influences semantic interpretation during sentence processing. To this aim, we manipulated sentence plausibility, rather than grammaticality, to investigate memory retrieval during language comprehension. Thus, while previous studies have examined illusions of grammaticality as evidence of retrieval interference, we probed for *illusions of plausibility* in fully grammatical, but implausible, sentences. We begin below by discussing cue-based parsing and illusions of grammaticality in more detail, before outlining previous research that has utilised plausibility effects to investigate the time-course of sentence processing.

Interference Effects in Language Comprehension

Successful language comprehension relies on the ability to form dependencies between non-adjacent constituents. For example, in (1a), there is a dependency between the verb 'read' and the non-adjacent constituent 'the book', which is interpreted as the verb's direct object even though it appears some words distant from the verb in the sentence. This type of dependency is typically called a *filler-gap* or *unbounded* dependency (Traxler & Pickering, 1996), as a displaced *filler* ('the book') needs to be associated with a corresponding *gap* (adjacent to 'read') at a later point in the sentence.

(1a) John saw the book that the boy very happily read while on holiday.

(1b) John saw the book that the boy with the magazine very happily read while on holiday.

According to cue-based parsing (McElree, 2000; McElree, Foraker & Dyer, 2003; Lewis & Vasishth, 2005; Lewis et al., 2006), dependency resolution in sentences like (1) involves retrieving a representation of the filler from memory at the verb. This retrieval operation is hypothesised to be guided by a set of cues that are matched against all items in memory in parallel. Cues can be derived from the local syntactic context, and other inherent properties of sentence constituents. For example, the verb 'read' may cue retrieval of an item marked as a [+DIRECT OBJECT], a feature that can be derived from the local syntactic context. Cues can also potentially be derived from other sources, such as the lexical properties of nouns and verbs (Van Dyke, 2007; Van Dyke & McElree, 2006). For example, the verb 'read' may cue retrieval of items marked as being [+READABLE]. In (1a), this combination of syntactic and lexical cues will uniquely identify the intended *retrieval target* 'the book'. In (1b) however, a *distractor* constituent partially matches the cues to retrieval, as 'the magazine' is a plausible but ungrammatical direct object of 'read'. As retrieval involves matching cues against all items in memory in parallel, the possibility of similarity-based

interference arises when multiple items in memory partially match a set of retrieval cues. In this case, distractor constituents, such as 'the magazine' in (2b), may sometimes be retrieved.

One example of interference during language processing comes from subject-verb agreement, as in (2), from Wagers, Lau & Phillips (2009).

(2a) The key to the cell unsurprisingly was rusty from many years of disuse.

(2b) The key to the cells unsurprisingly was rusty from many years of disuse.

(2c) The key to the cells unsurprisingly were rusty from many years of disuse.

(2d) The key to the cell unsurprisingly were rusty from many years of disuse.

In (2), the verb ('was/were') cues retrieval of the sentence subject ('the key'), which is the head of the phrase 'the key to the cell/s'. In (2a,b) this retrieval target matches the number properties of the verb, while (2c,d) is ungrammatical as the plural form of the verb mismatches the number properties of the singular subject. The ungrammaticality in (2c,d) leads to longer reading times compared to (2a,b). The size of this grammaticality effect is attenuated in (2c), when the distractor ('the cells') matches the number of the verb. Cue-based parsing explains this *illusion of grammaticality* as resulting from a partial-match between the cues at retrieval and the items held in memory. In (2c), no item *fully* matches the verb's retrieval cues (e.g. [+HEAD], [+PLURAL]), as the intended target is [+HEAD] but [-PLURAL], while the distractor is [+PLURAL] but [-HEAD]. On some proportion of trials, the partially-matching distractor may become activated to the extent that it is retrieved, which in turn will lead to an attenuation of the grammaticality effect and an illusion of grammaticality. Following Jäger, Engelmann & Vasishth (2017) we will refer to this pattern of results as *facilitatory* interference, as reading times for ungrammatical sentences are attenuated in the presence of a partially-matching distractor. Interference in such cases is

typically restricted to ungrammatical sentences, where no item in memory fully matches the retrieval cues (Dillon, Mishler, Slogett & Phillips, 2013; Lago, Shalom, Sigman, Lau & Phillips, 2015; Wagers et al., 2009). In addition to subject-verb agreement, facilitatory interference has also been reported for other linguistic dependencies (e.g. Parker & Phillips, 2016, 2017; Vasishth, Brüssow, Lewis & Drenhaus, 2008; Xiang, Dillon & Phillips, 2009; for review, see Jäger et al., 2017). To our knowledge however, facilitatory interference has not yet been observed in filler-gap dependencies.

Another type of interference has been reported in grammatical sentences. Van Dyke (2007) examined sentences such as (3), where successful comprehension requires retrieval of the sentence subject ('the worker') at 'was complaining'. However, a linearly closer distractor ('the warehouse'/'the neighbour'), which is an ungrammatical subject for this verb, intervenes. Van Dyke observed longer reading times after the critical verb when the distractor was a plausible ('the neighbour') compared to implausible ('the warehouse') subject for this verb, which was interpreted as indexing reanalysis following an initial misretrieval of plausible distractors on some proportion of trials. Following Jäger et al. (2017), we refer to this as *inhibitory* interference, as reading times are longer in grammatical sentences when a distractor partially matches the cues to retrieval.

(3) The worker was surprised that the resident who said that the warehouse/neighbour was dangerous was complaining about the investigation.

Similar results were reported by Van Dyke and McElree (2011) and by Glaser, Martin, Van Dyke, Hamilton and Tan (2013) using fMRI. Although inhibitory interference for some linguistic dependencies is thus well attested, for some other dependencies inhibitory interference has not been consistently reported. For example, in subject-verb agreement,

inhibitory interference might be expected in grammatical sentences, such that (2a), where both the retrieval target and distractor match the number properties of the verb, should have longer reading times compared to (2b), when only the retrieval target matches the verb's number. However, a number of studies have not reported such effects (Dillon et al., 2013; Lago et al., 2015; Wagers et al., 2009). Similarly in anaphora resolution, although one earlier study reported inhibitory interference in the resolution of reflexives and object pronouns (Badecker & Straub, 2002), subsequent studies have not consistently replicated this finding (Chow, Lewis & Phillips, 2014; Cunnings & Sturt, 2014; Dillon et al., 2013; Patterson, Trompelt, & Felser, 2014; Sturt, 2003). This may indicate that inhibitory interference effects are small and may be difficult to observe under certain conditions (for review, see Jäger et al., 2017). Indeed, in a recent study, Nicenboim, Vasishth, Engelmann and Suckow (2018) estimated inhibitory interference in subject-verb agreement to have a magnitude of 9ms, based on an analysis of 184 participants. As Nicenboim et al. argue, inhibitory effects may thus sometimes be numerically small and difficult to detect without adequate statistical power.

Although facilitatory and inhibitory interference effects have been taken as evidence for cue-based parsing, different implementations of cue-based retrieval account for such effects in different ways (for discussion, see Nicenboim & Vasishth, 2018). In the Lewis and Vasishth (2005) activation-based implementation, retrieval is a race process in which the item that receives the most activation is subsequently retrieved. In this model, the more active an item is, the faster it is retrieved. In this model, inhibitory interference in grammatical sentences indexes longer retrieval times as a result of activation spreading to distractor constituents, while facilitatory interference is explained in terms of the distractor being retrieved quicker than the retrieval target some proportion of the time, when the target doesn't fully match the cues to retrieval. In McElree's (2000) direct access account of cue-

based parsing however, retrieval speed is constant but the probability of retrieving the target representation is dependent on how well the cues match it and other distractors. In this way, differences in reading times relate to the probability of retrieving either the target or distractor constituent, rather than the speed of retrieval. Our study was not designed to tease apart these different accounts of similarity-based interference effects, and as such we do not discuss these two different implementations of cue-based parsing in detail. Instead, our focus is on how interference influences the semantic interpretation that is assigned to a sentence.

While illusions of grammaticality indicate that similarity-based interference can influence perception of sentence grammaticality, less is known about how such effects influence semantic interpretation. For example, in (2c), it is not known if readers misinterpret 'the cells' to be the subject of the predicate 'were rusty' rather than the grammatical sentence subject ('the key'). Some studies have utilised offline comprehension question measures as an index of how interference may influence the interpretation assigned to a sentence. Using forced-choice comprehension questions following word-by-word sentence reading, Patson and Husband (2016) found that readers sometimes misinterpret the number properties of nouns used in sentences like (2c), incorrectly answering 'yes' to questions such as 'Was there more than one key?'. Other researchers have also reported evidence of misinterpretation in other linguistic dependencies (e.g. Sturt, 2003; Van Dyke, 2007), but again these findings have been restricted to offline comprehension question measures. While such results provide some support that retrieval interference influences the interpretation assigned to sentences using offline measures, these effects may index interference during the post-trial question phase, rather than implicating interference during online processing as a sentence unfolds in real-time.

Plausibility Effects in Filler-Gap Dependencies

We aimed to investigate how interference influences semantic interpretation during online sentence processing. To this aim, we investigated how interference affects the reader's perception of semantic plausibility, as reflected in fixation times during reading. A number of existing studies have manipulated sentence plausibility to investigate the time-course of sentence processing during the resolution of direct-object filler-gap dependencies as in (1). Existing research in this vein has primarily examined the question of what conditions influence *when* the parser attempts to postulate a gap and construct a dependency. Many studies have shown that dependency formation in such cases is an *active* process, with comprehenders attempting to keep dependencies as short as possible (e.g. Frazier & Clifton 1989; Stowe, 1986; Traxler & Pickering, 1996). For example, Traxler and Pickering (1996) used a plausibility diagnostic to examine the time-course of filler-gap dependencies in sentences like (4).

(4a) We like the book that the author wrote unceasingly and with great dedication about while waiting for a contract.

(4b) We like the city that the author wrote unceasingly and with great dedication about while waiting for a contract.

(4c) We like the book/the city that the author who wrote unceasingly and with great dedication saw while waiting for a contract.

In (4a,b), the gap is ultimately at *about*, where both sentences are equally plausible, but there is an earlier possible gap at *wrote*, where (4a) is plausible but (4b) is implausible. Traxler and Pickering observed longer reading times at and after this verb in implausible condition (4b) compared to the plausible condition (4a), indicating readers postulated a gap

and attempted retrieval of the direct object filler ('the book'/'the city') at this first available possibility.[1]

Syntactic constraints may restrict when such dependencies can be formed however, and a number of studies have investigated how memory retrieval operations may be restricted during the resolution of filler-gap dependencies. To date, extant research investigating this issue has focused on so-called island constraints (for review, see Phillips, 2006). In addition to (4a,b), Traxler and Pickering also tested sentences like (4c), where a dependency at *wrote* is ungrammatical as a result of the relative clause island constraint, which restricts dependency formation between a filler and a verb inside a relative clause. In contrast to the plausibility effects observed in (4a,b), Traxler and Pickering observed similar reading times at and after *wrote* in (4c) for either filler. These results suggest readers did not postulate a gap inside a relative clause and thus did not attempt retrieval of the direct object filler at 'wrote' in (4c). These results suggest that syntactic island constraints restrict when retrievals may be attempted during the resolution of filler-gap dependencies (see also Felser, Cunnings, Batterham & Clahsen, 2012; Omaki & Schulz, 2011).

While studies on island constraints indicate that syntactic restrictions on *when* a memory retrieval operation can be initiated are respected during sentence processing, they do not inform our understanding about how different cues to memory retrieval may restrict *what* information is accessed once a retrieval operation is attempted. Results from studies such as Traxler and Pickering indicate that some representation of the filler must have been accessed

[1] Some accounts of active-gap filling assume that the filler is actively maintained in memory, rather than retrieved, at the gap site. Based on results examining filled-gap and plausibility effects in self-paced reading, Wagers and Phillips (2014) argue that while coarse-grained syntactic information may be actively maintained, allowing for active postulation of syntactic gaps, semantic information is indeed retrieved from memory at the gap site.

from memory at the critical verb in (4), but do not test how this retrieval operation itself is constrained, or whether it is susceptible to interference.

One study that bears on this issue was reported by Van Dyke and McElree 2006 (see also Van Dyke, Johns & Kukona, 2014). Van Dyke and McElree tested sentences like (5), where the filler ('the boat') needs to be retrieved at the verb 'sailed' in (5a) and 'fixed' in (5b). Participants read these sentences in a 'memory load' condition, where they also had to remember three words while reading of the sentence (e.g. table, sink, truck), and in a 'no memory load' condition where they did not have to remember any additional information. Crucially, the words in the memory load condition were plausible direct objects of the verb in (5b) but not (5a). Van Dyke and McElree reported longer reading times at the verb in (5b) compared to (5a) in the memory load condition but no differences were observed in the no memory load condition. This inhibitory interference effect suggests the items in the memory load condition caused retrieval interference at the critical verb.

(5a) It was the boat that guy who lived by the sea sailed in two sunny days.

(5b) It was the boat that the guy who lived by the sea fixed in two sunny days.

Instead of using a memory load manipulation, we utilised plausibility effects to investigate memory retrieval during the resolution of filler-gap dependencies. In addition to manipulating the plausibility of a retrieval target, as in previous studies such as Traxler and Pickering (1996), we additionally manipulated the plausibility of a distractor item in memory to examine the extent to which distractors influence the size of plausibility effects during the resolution of filler-gap dependencies. Thus, while previous studies have examined *illusions of grammaticality* as evidence of interference during sentence processing, we examined *illusions of plausibility* as evidence of interference during processing in otherwise

grammatical sentences. Below, we report two experiments that tested for illusions of plausibility during reading.

Experiment 1

In Experiment 1, we monitored participants' eye-movements as they read texts as in (6).

(6) The manor house was always very busy.

(a) *Plausible Sentence, Plausible Distractor*

Sue remembered the plate that the butler with the cup accidently shattered today in the dining room.

(b) *Plausible Sentence, Implausible Distractor*

Sue remembered the plate that the butler with the tie accidently shattered today in the dining room.

(c) *Implausible Sentence, Plausible Distractor*

Sue remembered the letter that the butler with the cup accidently shattered today in the dining room.

(d) *Implausible Sentence, Implausible Distractor*

Sue remembered the letter that the butler with the tie accidently shattered today in the dining room.

The owner of the house was not impressed.

In (6), the critical verb ‘shattered’ triggers retrieval of a constituent interpreted as its direct object.[2] The only *grammatical* constituent that can be retrieved at this verb is ‘the plate’ in (6a,b) and ‘the letter’ in (6c,d). In (6a,b) this grammatical retrieval target is a plausible direct object of this verb, while in (6c,d) it is implausible. We also manipulated the plausibility of an intervening distractor. In (6a,c) the distractor (‘the cup’) is a plausible, but ungrammatical, object of ‘shattered’, while in (6b,d) the distractor (‘the tie’) is implausible.

We expected longer reading times at the critical verb for implausible (6c,d) than plausible sentences (6a,b) (Traxler & Pickering, 1996). If retrieval interference influences semantic interpretation, this main effect of plausibility should be modulated by the distractor. If we observe an illusion of plausibility, reading times should indicate a pattern of facilitatory interference. In this case, reading times should be reliably shorter in (6c), when the distractor is plausible, compared to (6d), when it is implausible. We might also observe inhibitory interference, in which case we would expect longer reading times (6a), when both the retrieval target and distractor are plausible direct objects of the critical verb, compared to (6b), when only the target is plausible. Alternatively, if retrieval interference has no effect on the semantic interpretation assigned to a sentence, we should observe main effects of plausibility only in the absence of any effects of the distractor.

Participants

48 native English speakers from the University of Reading took part for course credit or were paid a nominal sum. All participants had normal or corrected-to-normal vision.

[2] A reviewer notes that the verb may trigger retrieval of both a subject and an object constituent, as both are nonadjacent to it. It is true that the verb potentially triggers two retrievals, but here we focus on retrieval of the object. Future research is required to investigate plausibility effects, and potential illusions of plausibility, during subject retrieval.

Materials

An initial set of 40 items was constructed as in (6). All items contained an introduction sentence, critical sentence and wrap-up sentence. We ran two pre-tests, each completed by 24 native English speakers who did not take part in the main eye-movement experiment, to ensure the items displayed the intended range of plausibility.

The first pre-test was a plausibility judgement questionnaire. The 40 items appeared in four conditions as in (7), which tested the plausibility of each retrieval target and distractor as a direct object of the critical verb. The 40 items (including introduction and wrap-up sentences) were pseudo-randomised across four lists in a Latin-square design with 40 fillers. Forward and reverse orders of each list were completed by the same number of participants, who rated the plausibility of each text on a scale from 1 (highly implausible) to 7 (highly plausible).

(7a) Sue remembered the plate that the butler accidently shattered in the dining room.

(7b) Sue remembered the cup that the butler accidently shattered in the dining room.

(7c) Sue remembered the letter that the butler accidently shattered in the dining room.

(7d) Sue remembered the tie that the butler accidently shattered in the dining room.

In order to rule out any unintended plausibility differences in the material immediately preceding the critical verb, a second pre-test rated the plausibility of the noun phrase that contained the manipulated distractor (‘the butler with the cup’, ‘the butler with the tie’). This was conducted as a ‘description’ rating study where participants had to rate the plausibility of each noun phrase, again on a scale from 1 (highly implausible) to 7 (highly plausible). The 40 critical items were mixed with 40 fillers across two lists in a Latin-square

design. Forward and reverse orders of this pre-test were again completed by the same number of participants.

Based on the results of the pre-tests, 32 items were used in the eye-movement experiment (see Appendix B). For these 32 items, results from the plausibility pre-test showed that sentences containing the plausible retrieval target (7a) and plausible distractor (7b) received similarly high mean plausibility ratings (for plausible targets, mean = 6.21, 95% CI = 6.00, 6.42; for plausible distractors, mean = 6.35, 95% CI = 6.19, 6.51). Sentences containing the implausible target (7c) and implausible distractor (7d) had similarly low ratings (for implausible targets, mean = 1.80, 95% CI = 1.74, 2.06; for implausible distractors, mean = 1.81, 95% CI = 1.53, 2.08). Results from the 'description' rating study indicated that these 32 items had similarly high ratings in the plausible distractor (mean = 6.45, 95% CI = 6.32, 6.58) and implausible distractor conditions (mean = 6.43, 95% CI = 6.26, 6.60).

For the main experiment, the 32 experimental items were mixed with 96 fillers that contained a variety of different syntactic structures. 16 of the fillers were ungrammatical as part of a manipulation from another study. All other fillers were grammatical and plausible. Comprehension questions that required a yes-no push-button response, which did not probe interpretation of the critical dependency, were asked after all critical trials and two thirds of the fillers.

Procedure

Items were pseudo-randomised across four lists in a Latin-square design. A different order was presented to each participant. Eye movements were recorded with an EYELINK 2000 sampling at 1000Hz. While viewing was binocular, eye-movements were recorded from the right eye. Each session began with calibration on a nine-point grid. Recalibration between

trials was conducted if required. Before each trial, participants fixated on a marker above the first word of the upcoming trial. Upon fixation on this marker, the text appeared. Participants read each text silently, pressing a button on a control pad once completed. The experiment lasted 30-45 minutes, and began with six practice trials.

Data Analysis

We report three reading time measures. Two are measures of first-pass processing, and one is a global index of processing that includes second-pass reading. For first-pass processing, we report *first pass times*, the sum of fixations within a region when it is first entered from the left, and *regression path times*, which sum fixations, starting when a region is first entered from the left, up until it is exited to the right. There is some debate in the literature regarding whether regression path time constitutes an 'early' or 'late' measure, but even if regression path times include some amount of processing incurred after a region is first exited to the left, it must index difficulty that was recognised during the first-pass of a particular region, before subsequent material is fixated. As such, we believe it provides the most complete measure of first-pass processing (Clifton, Traxler, Mohamed, Williams, Morris & Rayner, 2003). For a global index of processing, we report *total viewing times*, the total time spent in a region of interest, which sums both first- and second-pass processing, and, unlike regression path time, can include fixations after subsequent regions of text have been encountered.

We report these three measures at two regions of text. The *verb region* consisted of the critical verb that triggered retrieval, while the *spillover region* consisted of the following three words. In an attempt to minimise the number of separate statistical tests conducted on eye-movement data (von der Malsburg & Angele, 2017), and in contrast to previous analyses in which reading time measures are reported separately for each region, we report a single analysis for each eye-movement measure including 'region' as a fixed effect. Although this

does not eliminate issues related to multiple comparisons, it does help minimise the number of separate tests conducted across regions. We also believe this method provides two additional benefits over analysing regions separately. Firstly, it increases power to observe small effects that may be non-significant at individual regions but which nevertheless are consistent across regions. Secondly, many psycholinguistic studies, including our own previous work, make claims about the time-course of processing based on an effect being observed at a particular region but not another. For example, we may observe a particular effect in regression path times at the critical region and a different pattern of results in this same measure at the spillover region. Previous studies have typically not formally analysed these time-course effects statistically, even if conclusions about time-course are made. Including region as a fixed effect in the analysis formally tests the reliability of any potential time-course effects between regions.

Short fixations of 80ms or below within one degree of visual arc of another fixation were merged. All other fixations of 80ms or below, as well as those above 800ms, were removed before analysis. Trials in which a region was skipped were treated as missing data for that region. First-pass and regression path times at the critical verb were calculated using a leftward-shifting procedure of 4 characters to minimise skipping rates (see Sturt, 2003: 548). If the verb region was not fixated during the first-pass, fixations within 4 characters to the left of the verb were included as a fixation on the verb. Trials with excessive track loss were also removed before analysis. This accounted for less than 0.1% of the data.

The analysis was conducted using linear mixed-effects models with crossed random effects for subjects and items (Baayen, Davidson, & Bates, 2008). Analysis was conducted on log-transformed reading times to minimise skew (see Vasishth & Nicenboim, 2016). Models included sum coded (-1, 1) fixed main effects of ‘plausibility’ (plausible sentence vs. implausible sentence) ‘distractor’ (plausible distractor vs. implausible distractor) and region

(verb region vs. spillover region), along with their interactions. Each model was fit using the 'maximal' random effects structure (Barr, Levy, Scheepers, & Tily, 2013) that converged.[3] The maximal model included by-subject and by-item random intercepts, and random slopes for all fixed effects. By including 'region' as a fixed effect, our analysis includes two non-independent datapoints from each trial. To account for this, we included a random intercept for 'trial', defined as the unique subject and item pairing that constituted a particular trial. As the factor 'region' is the only repeated measure at the level of the trial, 'region' was the only random slope included under 'trial' (see Barr, 2013).

Following the recommendations of Luke (2017), *p* values were calculated using the Satterthwaite approximation. In the case of sentence plausibility by distractor interactions, planned comparisons tested for interference effects at the two levels of sentence plausibility. Below we do not interpret main effects of region, given that the critical verb and spillover regions have different lexical material, but interactions between region and the other manipulated variables provide insight into the time-course of processing. The full dataset and analysis code for both experiments reported here can be found at the first author's OSF website (https://osf.io/fe4us/).

Results

Mean comprehension question accuracy was 89% (all above 76%), indicating participants paid attention to the content of the sentences. Skipping rates during first-pass reading at the

[3] In case of non-convergence, we removed the random correlation parameters and refit the model (see Barr et al., 2013: 276). If this model still didn't converge, we iteratively removed the random correlation parameters that accounted for the least amount of variance in the data until convergence was achieved. For the main analyses reported for Experiments 1 and 2, removing the random correlation parameters always led to convergence.

verb and spillover regions were 6% and 3% respectively. Summaries of the eye-movement data and statistical analysis are provided in Tables 1 and 2.

***** INSERT TABLE 1 HERE *****

First pass and regression path times both revealed significant main effects of plausibility, with longer reading times in implausible than plausible sentences. There was also a significant interaction between region and plausibility in both measures, suggesting the size of the plausibility effect differed across the verb and spillover regions. Indeed, first pass times were on average 35ms longer, and regression path times 185ms longer, for implausible than plausible sentences at the spillover region. At the verb however, this difference was only 4ms in first pass times and 27ms in regression path times. No other main effects or interactions were significant in either measure.

***** INSERT TABLE 2 HERE *****

In total viewing times, we again observed a significant main effect of plausibility, with longer reading times in implausible sentences, and a significant region by plausibility interaction. As in regression path times, plausibility effects were larger at the spillover region than the verb region. We also observed a significant main effect of distractor and a significant plausibility by distractor interaction, in the absence of any further significant interactions with region. Nested contrasts were conducted on the data collapsed across region examining for effects of the distractor at the two levels of sentence plausibility. These indicated that there were no significant differences in total reading times in plausible sentences as a result of the plausibility of the distractor (estimate = 0.008, SE = 0.018, t = 0.44, p = .663), but total

viewing times for implausible sentences were significantly shorter when the distractor was plausible compared to implausible (estimate = 0.062, SE = 0.019, t = 3.28, p = .002). This pattern is illustrated in Figure 1, which shows total viewing times at the verb and spillover region. These results indicate an illusion of plausibility, with shorter reading times for implausible sentences when the distractor was a plausible, but ungrammatical, direct object of the critical verb.

****** INSERT FIGURE 1 HERE *****

Discussion

The results of Experiment 1 indicate an illusion of plausibility during online sentence processing. Total viewing times at the verb and spillover region were significantly longer for implausible than plausible sentences. This plausibility effect was however reliably attenuated when the distractor was a plausible, but ungrammatical, direct object of the critical verb. This pattern of faciliatation, indicating an illusion of plausibility, was analogous to the illusions of grammaticality that have been previously reported for subject-verb agreement (Dillon et al., 2013; Wagers et al., 2009). Our results suggest that the plausible distractor was retrieved on some proportion of trials, and indicates that retrieval interference can influence the ongoing semantic interpretation that is assigned to a sentence as it unfolds.

Although we observed facilitatory interference in implausible sentences, we did not observe significant effects of inhibitory interference in plausible sentences. We also observed main effects of plausibility in regression path times that were not significantly affected by the distractor, potentially suggesting a difference in time-course of plausibility and interference effects. The size of plausibility effects also differed across analysis regions. We discuss how

this pattern of results may have been obtained based on the time-course of processing, and the cues utilised during retrieval, in more detail in the General Discussion.

Having observed illusions of plausibility in Experiment 1, Experiment 2 aimed to replicate these findings. In particular, we manipulated the structure of the critical sentence to examine whether linguistic focus influences how susceptible memory retrieval operations are to interference.

Experiment 2

Focus is a linguistic device that can be used to add emphasis to a particular constituent of a sentence. For example, in (8b), the pseudo-cleft construction places focus on 'what' was bought compared to the standard active sentence (8a).

(8a) John bought the book.

(8b) What John bought was the book.

Focus is known to influence how constituents are processed during comprehension. For example, readers are more likely to correctly reject a statement as false if it is in focus (Bredart & Modolo, 1988), words are remembered more accurately when placed into focus (Birch & Garnsey, 1995), and readers are able to more accurately detect differences between two similar sentences in a change detection paradigm if the information that differs is focused (Sturt, Sanford, Stewart & Dawydiak, 2004). These findings suggest focused information receives an enhanced representation in memory. Foraker and McElree (2007) examined how focus influences memory representations in a series of experiments using the speed-accuracy trade-off paradigm. They found that focusing did not influence the speed at which information was accessed from memory, but focused information was more accurately

retrieved. They concluded that focused constituents receive more distinct representations in memory.

Based on these findings, testing whether we can replicate the illusions of plausibility observed in Experiment 1 in constructions where the retrieval target is placed into linguistic focus provides a particularly strong test of the generalisability of the interference effects we observed in Experiment 1. In Experiment 2, we thus tested sentences as in (9).

(9) The manor house was always very busy.

(a) *Plausible Sentence, Plausible Distractor*

What Sue remembered was the plate that the butler with the cup accidently shattered today in the dining room.

(b) *Plausible Sentence, Implausible Distractor*

What Sue remembered was the plate that the butler with the tie accidently shattered today in the dining room.

(c) *Implausible Sentence, Plausible Distractor*

What Sue remembered was the letter that the butler with the cup accidently shattered today in the dining room.

(d) *Implausible Sentence, Implausible Distractor*

What Sue remembered was the letter that the butler with the tie accidently shattered today in the dining room.

The owner of the house was not impressed.

The texts in (9) are identical to Experiment 1, except that the retrieval target ('the plate'/'the letter') now appears in a pseudo-cleft construction that places it into focus. As in Experiment 1, (9a,b) are plausible while (9c,d) are implausible. As such, we expected longer reading times at the critical verb 'shattered' in (9c,d) compared to (9a,b). If the illusions of plausibility observed in Experiment 1 replicate to conditions when the retrieval target is in focus, we should observe an attenuation of the plausibility effect when the distractor is plausible, leading to shorter reading times in (9c) than (9d). Alternatively, if items in focus receive more distinct representations in memory, linguistic focus may attenuate how susceptible items in memory are to interference. The strongest evidence that focus affects interference would be from an elimination of the interference effects observed in Experiment 1. In this case, we may observe significant main effects of plausibility only, in the absence of any significant effects of the distractor, in Experiment 2.

Participants

48 native English speakers from the University of Reading, none of whom took part in Experiment 1, took part in Experiment 2 for course credit or a nominal sum.

Materials

The 32 experimental items from Experiment 1 were adapted as in (9). All items were identical to Experiment 1 except that the retrieval target was now placed into focus in a pseudo-cleft construction. The experimental items were again interspersed with 96 fillers.

Procedure and Data Analysis

The procedure and data analysis was the same as in Experiment 1. Less than 0.1% of trials were removed due to excessive track loss.

Results

Comprehension question accuracy was 94% (all above 87%). Skipping rates during first-pass processing were 7% and 3% for the verb and spillover regions respectively. Summaries of the eye-movement data and statistical analysis are provided in Tables 3 and 4.

***** INSERT TABLE 3 HERE *****

Analysis of first pass times indicated a significant main effect of plausibility, with longer reading times for implausible than plausible sentences. In regression path times, there was again a significant main effect of plausibility, with longer reading times in implausible sentences, that was this time modulated by an interaction with region. As in Experiment 1, the plausibility effect was numerically larger at the spillover region than the verb region. Regression path times also revealed a significant plausibility by distractor interaction, in the absence of any other interactions with region. Nested contrasts, collapsed across regions, indicated no significant differences in regression path times for plausible sentences (estimate = -0.016, SE = 0.019, t = -0.82, p = .411). For implausible sentences however, regression path times were significantly shorter when the distractor was plausible compared to implausible (estimate = 0.040, SE = 0.020, t = 2.01, p = .047), indicating an illusion of plausibility.

***** INSERT TABLE 4 HERE *****

For total viewing times, there was again a main effect of plausibility with longer reading times in implausible sentences. The main effect of distractor was significant, as was the plausibility by distractor interaction, in the absence of any reliable interactions with

region. Nested contrasts indicated that total viewing times for plausible sentences did not differ significantly as a result of the plausibility of the distractor (estimate = -0.001, SE = 0.017, t = -0.07, p = .948), but total viewing times for implausible sentences were significantly shorter when the distractor was plausible compared to implausible (estimate = 0.053, SE = 0.018, t = 2.98, p = .005). This pattern of results, which is very similar to those observed in total viewing times in Experiment 1, is shown in Figure 2. As in Experiment 1, these results indicate an illusion of plausibility.

***** INSERT FIGURE 2 HERE *****

Discussion

The main finding of Experiment 2 was a replication of the illusion of plausibility from Experiment 1 in total viewing times. At the critical verb and spillover region, the size of the plausibility effect was reliably attenuated when the distractor was a plausible, but ungrammatical, direct object of the verb. Another similarity between Experiments 1 and 2 was that the plausibility effect was significantly larger at the spillover region than the verb region in regression path times. One difference between Experiments 1 and 2 however was that we also found evidence of an illusion of plausibility in regression path times in Experiment 2 but not Experiment 1. We discuss these findings in more detail below.

In sum, the results of Experiment 2 suggest that focused items are susceptible to retrieval interference, as was also the case for non-focused items in Experiment 1. We discuss the implications of our findings from both experiments in more detail below.

General Discussion

The main aim of this study was to investigate whether retrieval interference influences the semantic interpretation that is assigned to a sentence as it unfolds during processing. We also tested if linguistic focus alters the extent to which an item in memory is susceptible to interference. Our results also bear on questions related to the nature of the set of retrieval cues that may be implemented during language comprehension. We discuss each of these issues in turn below.

Retrieval Interference and Semantic Interpretation

Our results provide clear evidence that retrieval interference affects the unfolding semantic interpretation that is assigned to a sentence as it is processed. Facilitatory interference was most clearly observed in both experiments in total viewing times, where the size of plausibility effects was reliably attenuated when the distractor was a plausible, but ungrammatical, direct object of the critical verb. These illusions of plausibility provide strong evidence that retrieval interference influences semantic interpretation. Existing studies testing how interference influences the interpretation assigned to a sentence required explicit responses to post-trial questions in comparatively 'offline' measures (e.g. Patson & Husband, 2015; Sturt, 2003; Van Dyke, 2007), rather than investigating such effects during online language processing. Furthermore, while studies investigating *illusions of grammaticality* have shown that interference can influence perception of sentence grammaticality during processing, such studies have not shown how such effects influence semantic interpretation. The *illusions of plausibility* reported here indicate that retrieval interference can have a potent effect on semantic processing. Similar to illusions of grammaticality, we observed illusions of plausibility in sentences in which no item fully matches a set of retrieval cues. In both cases, the retrieval target does not fully match the cues at retrieval. However, while existing studies of facilitatory interference have been restricted to sentences that are ungrammatical, the

illusions of plausibility here implicate facilitatory interference in grammatical sentences, where an item in memory matches any grammatical cues that may guide retrieval.

In both experiments, plausibility effects were observed in first pass times, while illusions of plausibility were observed in total viewing times. In Experiment 1 we also observed plausibility effects only, but not interference, in regression path times, while we did find evidence of interference in this measure in Experiment 2. Some have claimed that an initial stage of retrieval during sentence processing targets grammaticality licit constituents only, with grammatically illicit distractors having a delayed effect that follows retrieval of the grammatical target constituent (e.g. Cunnings & Sturt, 2014; Sturt, 2003). A similar argument, that facilitatory interference reflects reanalysis, has also been claimed based on the time-course of effects in subject-verb agreement attraction (Lago et al, 2015; Wagers et al., 2009; but see Tanner, Grey & Van Hell, 2017, for contrasting evidence). While the observed differences in, for example, first-pass times and total viewing times in our experiments might be compatible with this time-course of effects, we note that there was some evidence of interference fairly early, in regression path times, in Experiment 2. We are thus cautious in drawing any strong conclusions about the time-course of interference effects here.

Although we found clear evidence for facilitatory interference, we did not observe significant effects of inhibitory interference in either experiment. While this parallels work on illusions of grammaticality (e.g. Wagers et al., 2009), we are cautious in interpreting null effects as evidence for the absence of an effect, and do not make any strong claims about inhibitory interference in plausible sentences. Indeed, inhibitory interference has been clearly observed in other studies that examined other syntactic configurations (e.g. Van Dyke, 2007; Van Dyke & McElree, 2011) or utilised a memory load paradigm (Van Dyke & McElree, 2006). As illustrated by Nicenboim et al. (2018), it might be that our study did not have adequate statistical power to observe inhibitory interference effects, which may be smaller

than the facilitatory effects that we did observe. It might also be that our manipulation was not strong enough to induce inhibitory interference (see discussion in Van Dyke & McElree, 2011, on how distractors may need to match relevant syntactic features to cause inhibitory interference). Further research is required to investigate the extent to which different contexts engender inhibitory and facilitatory interference. Irrespective of these issues, we maintain that the facilitatory interference effects that we observed in both experiments are compatible with the hypothesis that distractors are sometimes retrieved, at least when the retrieval target does not fully match the set of retrieval cues.

Plausibility Effects, Interference and Linguistic Focus

In both experiments, especially in regression path times, plausibility effects were larger at the spillover region than the verb region. Note that the spillover region was larger than the verb region, consisting of three rather than a single word, and as such the larger plausibility effects at the spillover region may in part reflect that people in general spent longer reading this region than the single-word verb region. For example, if our experimental manipulation affected the durations of a series of fixations following the reader's first encounter with the critical word, a longer region might capture more of the affected fixations, and since eye-movement measures involve summing fixation durations, this could lead to a larger difference between conditions. However, as one reviewer notes, since standard deviations tend to increase with increasing mean response times (Wagenmakers & Brown, 2007), it might also be that the spillover region provided an exaggerated estimate of the size of plausibility effects. However the plausibility effects at the critical and spillover region are interpreted, our results indicate that analysis of any one region, even if it is critical to the analysis, may not provide a full estimate of how an experimental manipulation may influence reading times across larger portions of text.

The size of both plausibility effects and interference effects were overall quite similar in Experiments 1 and 2. Indeed, the pattern of results for total viewing times illustrated in Figures 1 and 2 are similar for both experiments. We reasoned that if focused constituents receive enhanced representations in memory (e.g. Foraker & McElree, 2007), then they may be less susceptible to interference than non-focused constituents. The strongest support for this would have been from an elimination of the illusion of plausibility that was observed in Experiment 1 in Experiment 2. The results of Experiment 2 clearly did not support this hypothesis, as we found clear evidence of interference. To statistically test the extent to which focusing may influence interference, we conducted additional post-hoc analyses of our data including 'experiment' as independent variable. In this analysis, we analysed the data from experiments 1 and 2 together, including 'experiment' as a sum-coded between-subject, but within-item, fixed effect.

For first pass times, this analysis revealed a significant main effect of plausibility (estimate = 0.023, SE = 0.006, t = 3.55, p = .001) in the absence of any other significant effects. For regression path times, this yielded significant main effects of plausibility (estimate = 0.081, SE = 0.011, t = 7.27, p < .001) and distractor (estimate = 0.019, SE = 0.009, t = 2.23, p = .034), a significant plausibility by region interaction (estimate = 0.057, SE = 0.010, t = 5.64, p < .001), and a significant 3-way experiment by plausibility by distractor interaction (estimate = 0.016, SE = 0.008, t = 1.98, p = .048). In the main analysis, recall that we observed a significant plausibility by distractor interaction in regression path times in Experiment 2, suggestive of facilitatory interference, but not in Experiment 1. Contrary to the hypothesis that focusing may reduce interference, if anything, this 3-way interaction suggests focused items in Experiment 2 were more susceptible to interference than the non-focused items in Experiment 1 in this measure. For total viewing times, there were significant main effects of plausibility (estimate = 0.110, SE = 0.009, t = 12.38, p < .001) and

distractor (estimate = 0.032, SE = 0.009, t = 3.51, p = .001), and a plausibility by region interaction (estimate = 0.016, SE = 0.007, t = 2.18, p = .037). Crucially, the plausibility by distractor interaction was also significant (estimate = 0.026, SE = 0.008, t = 3.38, p < .001), in the absence of any further significant main effects or interactions. Nested contrasts for the combined data indicated no significant effects of the distractor in plausible sentences (estimate = 0.003, SE = 0.012, t = 0.28, p = .777), but for implausible sentences, reading times were significantly shorter when the distractor was plausible (0.057, SE = 0.015, t = 3.93, p < .001).

In summary, we did not find evidence that focused constituents are less susceptible to interference than non-focused constituents. If anything, the time-course of results in regression path times suggest a trend in the opposite direction. We are cautious in interpreting our results in regression path times relating to the time-course of interference for focused and non-focused constituents until further investigation however, ideally using a within-subject manipulation. Indeed, it is clear from various studies that focus does influence how items are represented in memory (Birch & Garnsey, 1995; Bredart & Modolo, 1988; Foraker & McElree, 2007; Sturt et al., 2004). However, while we cannot conclude what specific effect focusing may have on memory encoding and retrieval during sentence processing, our results nevertheless indicate that focusing does not alter memory representations to the extent that focused constituents are impervious to interference. Both focused and non-focused constituents appear susceptible to retrieval interference during comprehension.

Retrieval Cues in Filler-Gap Dependencies

A final issue to consider relates to the set of retrieval cues that may have led to the interference effects that we observed. For agreement attraction, as in (2), retrieval interference is typically described in terms of an interaction between the syntactic (e.g.

[+HEAD]) and morphological ([+PLURAL]) features that are utilised as cues to guide retrieval. Van Dyke (2007) and Van Dyke and McElree (2006) argued that other retrieval cues can be drawn from the lexical properties of specific verbs. We believe that the interference effects that we observed are best described in terms of an interaction between such lexical retrieval cues and syntactic cues. Consider the retrieval cues that may be implemented during comprehension to guide retrieval in the critical sentences we tested, exemplified in (10).

(10) Sue remembered the plate/the letter that the butler with the cup/the tie accidently shattered today in the dining room.

When the retrieval target ('the plate'/'the letter') is first encountered during reading, based on the local syntactic context it can be encoded in memory with features denoting its syntactic role, such as [+DIRECT OBJECT].[4] When the critical verb, which is missing an overt object, is encountered, retrieval could contain cues for an item in memory marked as [+DIRECT OBJECT]. Other retrieval cues may be derived from the lexical properties of the verb that cues retrieval. However, while number cues (e.g. [+PLURAL]) as utilised in subject-verb agreement can be derived from nouns and verbs as they are encountered during processing, constituents cannot in any obvious way be encoded as being [+PLAUSIBLE] or [+IMPLAUSIBLE], given that there is nothing plausible or implausible about the retrieval target or distractor at the point in the sentence at which they are first encountered. As such,

[4] Retrieval here could equally rely on features based on the semantic representation, such as [+THEME]. For current purposes, either characterisation would lead to the same predictions. As it is not possible to distinguish between these two accounts with the current data, we do not discuss this issue further.

plausibility in itself cannot be utilised as a retrieval cue at the verb. Core lexical semantic distinctions, such as animacy, could be encoded on items in memory as a sentence is processed and then utilised as a retrieval cue at the verb, if certain verbs favour animate or inanimate arguments. However, although animacy may be a cue that guides retrieval, retrieval targets and distractors in our experiments were inanimate in both plausible and implausible conditions, and as such this cue would not be able to account for the pattern of interference that we found.

Instead, the illusions of plausibility that we observed are likely the result of *verb-specific* retrieval cues derived from the idiosyncratic properties of individual lexical items (see also Van Dyke & McElree, 2006). For example, in (10), we propose that the verb *shattered* triggers retrieval of items marked as being [+SHATTERABLE] (or some other feature, or set of features, that denote this semantic property). Nouns in the sentence would also need to be marked with relevant features. For example, in (10), 'the plate' and 'the cup' would need to be marked as [+SHATTERABLE] but not 'the letter' or 'the tie'. In plausible conditions, the combination of the [+DIRECT OBJECT] and [+SHATTERABLE] cues will uniquely identify the retrieval target, which will thus become most highly activated and most likely retrieved. In implausible conditions, the retrieval target matches the [+DIRECT OBJECT] feature but not the [+SHATTERABLE] feature, while the distractor does not match [+DIRECT OBJECT] but is [+SHATTERABLE]. As no item fully matches the retrieval cues, on some proportion of trials, the plausible distractor will thus become activated to the extent that it is retrieved, leading to an illusion of plausibility. For other verbs, other verb-specific cues would be utilised as retrieval cues in a similar fashion.

While this proposal would account for the pattern of results that we observed, it potentially greatly increases the number of features that need to be encoded on sentence chunks held in memory than is typically assumed in implemented models of cue-based

retrieval (e.g. Lewis & Vasishth, 2005). Features derived from core morphological or lexical properties (e.g. [+PLURAL], [+ANIMATE]) do not in themselves greatly increase the number of features that need to be encoded on items in memory, given that all nouns have number and animacy properties. But given that any number of idiosyncratic properties of a noun could be utilised as a cue during subsequent retrievals, and given that when a noun is first encountered the reader has little clue as to *which* idiosyncratic features may later cue retrieval, the list of potential features encoded on items in memory becomes increasingly large. While such features can be easily derived from the corresponding lexical representation, this approach may require an unfeasibly large number of features to be encoded on sentence chunks in memory. One possibility could be that only a subset of core semantic properties derived from lexical items are encoded in the relevant sentence chunks in memory, with other potential features that could be encoded being left unspecified. Only specified features could then be activated during subsequent retrievals. While this reduces the burden on the total number of features that need to be encoded on sentence chunks in memory, what constitutes a 'core' semantic feature is an issue that requires further empirical attention. Further research is required to examine how idiosyncratic lexical features may cue retrieval operations during language comprehension. However such cues are implemented, we maintain that the illusions of plausibility that we report result from *verb-specific* retrieval cues derived from the idiosyncratic properties of individual lexical items.

Conclusion

We investigated retrieval interference during sentence processing by manipulating sentence plausibility. The plausibility effects that we observed were crucially attenuated when a distractor constituent in memory was an ungrammatical, but plausible, retrieval target of a critical verb. This illusion of plausibility indicates that retrieval interference affects the

semantic interpretation that is assigned to a sentence as it is processed, and we believe is best described in terms of verb-specific retrieval cues that may guide retrieval to grammatically illicit, but plausible, constituents during the resolution of filler-gap dependencies. These illusions of plausibility indicate that interference has a potent influence on the semantic interpretation that is assigned to a sentence and provide a new diagnostic for investigating retrieval interference during language comprehension.

Acknowledgements

The research reported here was funded by a British Academy/Leverhulme Trust Small Research Grant. We would like to thank Evelyn Egger and Abhijeet Patra for their help with data collection.

Appendix A

Below are the experimental items from Experiment 1. Sentence plausibility and distractor plausibility manipulations are delimited with a forward slash (/). The experimental items in

Experiment 2 were the same, with the exception that the retrieval target was placed into focus, as exemplified in (9) in the main text.

1 The manor house was always very busy. Sue remembered the plate/letter that the butler with the cup/tie accidently shattered today in the dining room. The owner of the house was not impressed.

2 It was getting quite late in the forest. John noticed the tent/tree that the camper by the chair/river very quietly moved before going off to sleep. It was going to be a quiet night.

3 It was another beautiful day in the field. Jane saw the tree/tent that the lady by the flowers/fence thoughtfully planted a number of years earlier. There was time to appreciate everything that day.

4 The kitchen was a difficult place to work in. Jim saw the chicken/cutlery that the chef by the food/sink very expertly cooked despite all of the pressure. The diners were in for a treat.

5 The countryside was really pretty. Vicky noticed the corn/gate that the farmer by the crops/fence very recently harvested with a big new tractor. There was lots of work still to be done.

6 It's difficult to get home during rush hour. Pauline saw the van/box that the man by the car/bin very carefully drove despite all the bad traffic. It took hours to get back to the house.

7 A serious crime had been committed. Gregory remembered the diamond/house that the robber with the money/cough very secretly stole late at night last week. The police were on the case.

8 There was lots to do in the classroom. Jessica liked the poem/pen that the student with the script/pencil passionately recited a few hours ago today. It was a busy time at school.

9 It got quite cold in the wilderness. Matthew remembered the wood/pond that the man by the tree/lake quite expertly chopped to help build a fire. It really helped warm things up at night.

10 There was a lot to do at the company. Bernadette saw the email/laptop that the manager with the report/office very hastily wrote earlier in a real hurry. There were going to be some changes at work.

11 It was certainly a nice place to live. Mary noticed the garage/flower that the couple by the house/tree very cleverly built a couple of years earlier. It had been quite a task.

12 The summer was very warm that year. Nigel saw the daffodil/pavement that the guy near the rose/door very happily watered as the sun shone brightly. It really was very hot outside.

13 The supermarket was always very busy. Monica saw the trolley/building that the man by the cart/shop quite hastily pushed while looking for some food. Some people have very little patience.

14 The restaurant had some very good reviews. Julia saw the bread/beer that the lady with the meal/wine quite happily ate during an expensive night out. Sometimes good food can cost a lot.

15 Sometimes things happen by mistake. Warren noticed the glass/shirt that the girl by the tiles/food accidently cracked earlier near the kitchen sink. It wasn't really anybody's fault.

16 The lobby was starting to get busy. Hillary remembered the shirt/glass that the teenager with the hat/pen very happily wore during a big night out. A group of friends had gone to the cinema.

17 There are always lots of chores to do. Barry saw the table/potato that the housewife with the ornament/shopping carefully dusted an hour or two ago. It was boring but needed to be done.

18 Some children can be rather careless. Joanna noticed the water/spoon that the boy with the juice/napkin very stupidly poured during a break at school. It went absolutely everywhere.

19 Sometimes classes can be quite boring. Nathan noticed the note/bag that the student with the essay/ruler patiently wrote during a very dull class. The lesson seemed to last for hours.

20 It was a rather lazy day. Tabitha remembered the newspaper/candlestick that the lady by the magazines/fireplace very slowly read while relaxing back at home. There was little else to do.

21 Some kids never pay attention to others. Tim saw the skateboard/cardboard that the teenager by the bike/pond dangerously rode without looking where to go. A bad accident very nearly occurred.

22 The restaurant was quite well known. Annabelle saw the steak/knife that the cook with the pork/fork quite expertly grilled during the rush at lunch. It was always busy around noon.

23 It can be tense in the armed forces. Jake noticed the helicopter/automobile that the pilot by the plane/tank rather expertly landed a couple of nights ago. The weather had been terrible that evening.

24 The summer holiday had been fantastic. Wendy remembered the yacht/beach that the man with the boat/shirt quite confidently sailed one hot and sunny day. It was an enjoyable trip.

25 The music school was known around the world. Barry saw the piano/stage that the pupil by the guitar/curtain passionately played during a recital one day. It sounded really good.

26 There was an awful lot still to do. Jessica saw the cupcake/handbag that the lady by the bread/kettle very carefully baked while in the kitchen today. Some people were coming round for tea.

27 The weather was really good that afternoon. Bob noticed the orange/towel that the boy with the apple/juice really slowly peeled after relaxing in the sun. It was time for a quick snack.

28 It was a surprisingly cold winter. Janice admired the jumper/fireplace that the woman with the scarf/paper quite happily knitted while at home last week. It was important to keep warm.

29 Things can be hectic in the morning. Melissa liked the crumpet/teapot that the woman with the bread/plate very quickly toasted one busy day before work. There was barely enough time for breakfast.

30 A new mall had opened in the city centre. Jimmy noticed the money/store that the guy with the cash/cap needlessly spent during a large shopping spree. Some people don't know when to stop.

31 Everyone was excited about the Summer Ball. Sophia liked the dress/music that the girl with the coat/money quite happily wore to the party last weekend. Everybody had a great time.

32 It got extremely intense during the war. Vince found the rifle/bandage that the soldier near the gun/house carelessly fired before taking time to aim. It was a reckless thing to do.

References

Baayen, H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Barr, D. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328. https://doi.org/10.3389/fpsyg.2013.00328

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278.

Benjamin, S., Desmet, T., Clifton, C. and Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*, 201-216.

Birch, S. & Garnsey, S. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language*, *34*, 232-267.

Bredart, S. & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, *67*, 135-144.

Christianson, K., Hollingworth, A., Halliwell, J. & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368-407.

Clifton, C., Traxler, M., Mohamed, M., Williams, R., Morris, R. & Rayner, K. (2003). The use of thematic information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*, 317-334

Chow, W-Y., Lewis, S. & Phillips, C. (2014). Immediate sensitivity to structural constraints on pronoun resolution. *Frontiers in Psychology*, *5*, 630. https://doi.org/10.3389/fpsyg.2014.00630

Cunnings, I. and Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory and Language*, *75*, 117-139.

Dillon, B., Mishler, A., Sloggett, S. & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modelling evidence. *Journal of Memory and Language*, *69*, 85-103.

Felser, C., Cunnings, I., Batterham, C. & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, *34*, 67-98.

Foraker, S. & McElree, B. (2007). The role of prominence in pronoun resolution: Availability vs. accessibility. *Journal of Memory and Language*, *56*, 357-383.

Frazier, L. & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes, 2*, 93-126.

Glaser, Y., Martin, R., Van Dyke, J., Hamilton, A. & Tan, Y. (2013). Neural basis of semantic and syntactic interference in sentence comprehension. *Brain and Language*, *126*, 314-326.

Jäger, L., 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.

Jonides, J., Lewis, R., Nee, D., Lustig, C., Berman, M. and Moore, K. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*, 193-224.

Lewis, R. & Vasishth, S. (2006). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 375-419.

Lewis, R., Vasishth, S., and Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*, 44-54.

Luke, S. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*, 1494-1502.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*, 111-123.

McElree, B., Foraker, S. & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language, 48*, 67-91.

Omaki, A. & Schulz, B. (2011). Filler-gap dependencies and island constraints in second language processing. *Studies in Second Language Acquisition*, *33*, 563-588.

Nicenboim, B. and Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modelling. *Journal of Memory and Language*, *99*, 1-34.

Nicenboim, B., Vasishth, S., Engelmann, F. and Suckow, K. (2018). Exploratory and confirmatory analyses in sentence processing: A case study of number interference in German. *Cognitive Science*. Available online, https://doi.org/10.1111/cogs.12589

Parker, D. & Phillips, C. (2016). Negative polarity items and the format of hierarchical encodings in memory. *Cognition*, *157*, 321-339.

Parker, D. & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language*, *94*, 272-290.

Patson, N. & Husband, E. (2016). Misinterpretations in agreement and agreement attraction. *The Quarterly Journal of Experimental Psychology*, *69*, 950-971.

Patterson, C., Trompelt, C. & Felser, C. (2014). The online application of binding condition B in native and non-native pronoun resolution. *Frontiers in Psychology*, *5*, 147. https://doi.org/10.3389/fpsyg.2014.00147

Phillips, C. (2006). The real-time status of island phenomena. *Language*, *82*, 795-823.

Phillips, C., Wagers, M. & Lau, E. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In Runner, J. (ed.), *Experiments at the Interfaces*. *Syntax and Semantics, 37*, 153-186. Bingley, UK: Emerald Publications.

Stowe, L. (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*, 227-245.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*, 542-562.

Sturt, P., Sanford, A., Stewart, A. & Dawydiak, E. (2004). Linguistic focus and good enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, *11*, 882-887.

Tanner, D., Grey, S. & Van Hell, J. (2017). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology*, *54*, 248-259.

Traxler, M. & Pickering, M. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*, 454-476.

Van Dyke, J. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *33*. 407-430.

Van Dyke, J. & Johns, C. (2012). Memory interference as a determinant of language comprehension. *Language and Linguistics Compass*, *6*, 193-211.

Van Dyke, J., Johns, C. & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition*, *131*, 373-403.

Van Dyke, J. & McElree (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*, 157-166.

Van Dyke, J. & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*, 246-263.

Vasishth, S., Brüssow, S., Lewis, R. & Drenhaus, H (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*, 685-712.

Vasishth, S. & Nicenboim, B. (2016). Statistical methods for linguistic research: Foundational Ideas – Part I. *Language and Linguistics Compass*, *10*, 349-369.

von der Malsburg, T. & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, *94*, 119-133.

Wagenmakers, E-J., & Brown, S. (2007). On the linear relationship between the mean and the standard deviation of a response time distribution. *Psychological Review, 114,* 830-841.

Wagers, M., Lau, E. & Phillips, C. (2009). Attraction in comprehension: representation and processes. *Journal of Memory and Language, 61*, 206-237.

Wagers, M. & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *Quarterly Journal of Experimental Psychology, 67*, 1274-1304.

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain & Language, 108*, 40-55.

Table 1. *Summary of eye-movement measures in Experiment 1 (SDs in parentheses)*

	First Pass Time		Regression Path Time		Total Viewing Time	
Verb Region						
Plausible Sentence, Plausible Distractor	250	(102)	317	(275)	337	(205)
Plausible Sentence, Implausible Distractor	259	(110)	346	(451)	343	(189)
Implausible Sentence, Plausible Distractor	262	(111)	364	(416)	400	(314)
Implausible Sentence, Implausible Distractor	256	(102)	354	(362)	450	(299)
Spillover Region						
Plausible Sentence, Plausible Distractor	428	(266)	634	(722)	614	(389)
Plausible Sentence, Implausible Distractor	433	(261)	663	(712)	629	(434)
Implausible Sentence, Plausible Distractor	471	(294)	900	(1127)	777	(533)
Implausible Sentence, Implausible Distractor	463	(272)	949	(1117)	855	(501)

Table 2. *Summary of the statistical analysis in Experiment 1*

	Model Estimate (SE)	*t*	*p*
First Pass Time			
Region	0.230 (0.025)	9.35	< .001
Plausibility	0.026 (0.009)	2.87	.007
Distractor	0.003 (0.008)	0.43	.667
Region * Plausibility	0.018 (0.009)	2.05	.049
Region * Distractor	0.001 (0.010)	0.14	.894
Plausibility * Distractor	-0.011 (0.009)	-1.16	.256
Region * Plausibility * Distractor	0.002 (0.008)	0.27	.790
Regression Path Time			
Region	0.334 (0.031)	10.79	< .001
Plausibility	0.079 (0.015)	5.22	< .001
Distractor	0.025 (0.013)	1.94	.064
Region * Plausibility	0.053 (0.014)	3.89	< .001
Region * Distractor	0.011 (0.013)	0.86	.399
Plausibility * Distractor	-0.004 (0.011)	-0.33	.739
Region * Plausibility * Distractor	0.004 (0.011)	0.38	.703
Total Viewing Time			
Region	0.300 (0.028)	10.85	< .001
Plausibility	0.114 (0.012)	9.26	< .001
Distractor	0.036 (0.013)	2.85	.008
Region * Plausibility	0.022 (0.008)	2.73	.006
Region * Distractor	0.001 (0.008)	0.18	.860
Plausibility * Distractor	0.024 (0.011)	2.29	.022
Region * Plausibility * Distractor	0.003 (0.008)	0.33	.740

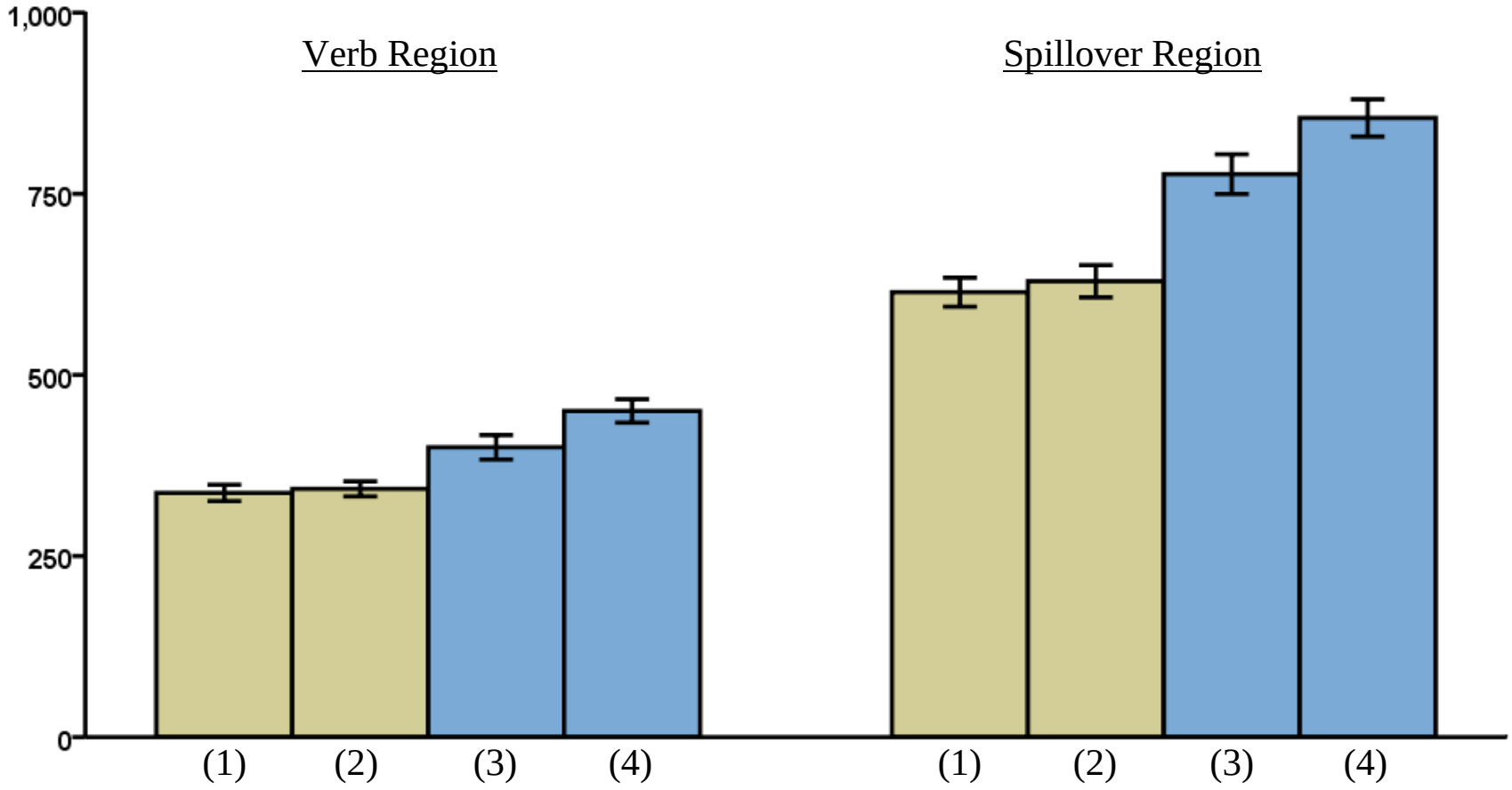
Table 3. *Summary of eye-movement measures in Experiment 2 (SDs in parentheses)*

	First Pass Time		Regression Path Time		Total Viewing Time	
Verb Region						
Plausible Sentence, Plausible Distractor	250	(96)	377	(527)	348	(216)
Plausible Sentence, Implausible Distractor	254	(109)	320	(300)	342	(204)
Implausible Sentence, Plausible Distractor	263	(104)	343	(341)	396	(249)
Implausible Sentence, Implausible Distractor	270	(146)	355	(305)	473	(316)
Spillover Region						
Plausible Sentence, Plausible Distractor	440	(280)	607	(607)	656	(454)
Plausible Sentence, Implausible Distractor	432	(270)	599	(602)	651	(435)
Implausible Sentence, Plausible Distractor	433	(265)	859	(1195)	781	(527)
Implausible Sentence, Implausible Distractor	475	(281)	1005	(1175)	878	(595)

Table 4. *Summary of the statistical analysis in Experiment 2*

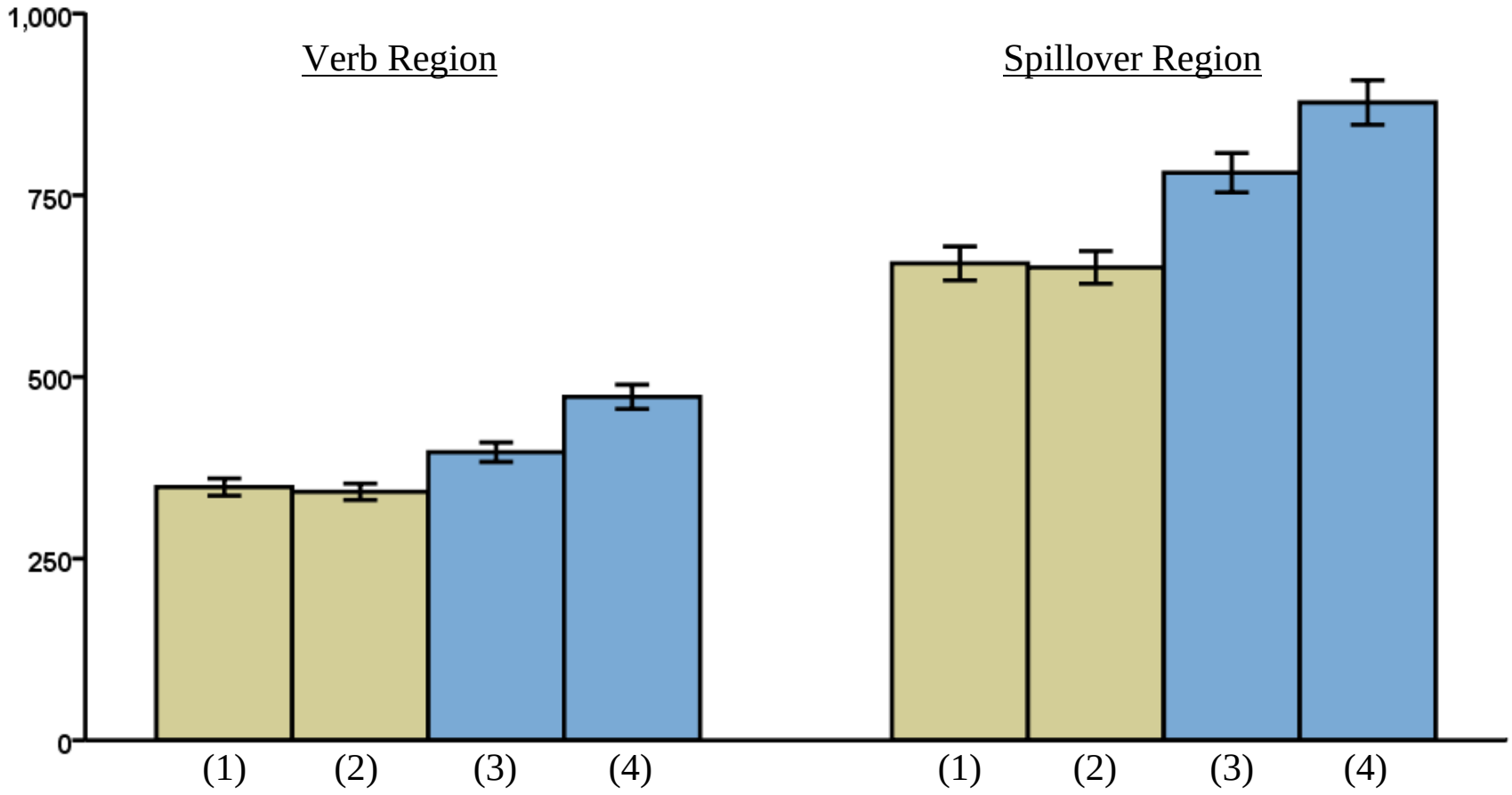
	Model Estimate (SE)	*t*	*p*
First Pass Time			
Region	0.222 (0.023)	9.64	< .001
Plausibility	0.019 (0.008)	2.35	.019
Distractor	0.009 (0.008)	1.10	.271
Region * Plausibility	0.001 (0.009)	0.06	.952
Region * Distractor	0.009 (0.009)	1.00	.330
Plausibility * Distractor	0.009 (0.010)	0.91	.369
Region * Plausibility * Distractor	0.014 (0.010)	1.40	.176
Regression Path Time			
Region	0.319 (0.030)	10.71	< .001
Plausibility	0.083 (0.015)	5.63	< .001
Distractor	0.014 (0.013)	1.12	.274
Region * Plausibility	0.061 (0.014)	4.37	< .001
Region * Distractor	0.025 (0.013)	1.85	.076
Plausibility * Distractor	0.028 (0.011)	2.46	.014
Region * Plausibility * Distractor	0.009 (0.013)	0.65	.523
Total Viewing Time			
Region	0.300 (0.027)	11.00	< .001
Plausibility	0.106 (0.013)	7.97	< .001
Distractor	0.028 (0.012)	2.34	.026
Region * Plausibility	0.009 (0.011)	0.85	.404
Region * Distractor	-0.002 (0.009)	-0.22	.826
Plausibility * Distractor	0.027 (0.011)	2.50	.012
Region * Plausibility * Distractor	-0.004 (0.009)	-0.46	.648

Figure 1. *Total Viewing Times in Milliseconds for Experiment 1*



(1) Plausible Sentence, Plausible Distractor
(2) Plausible Sentence, Implausible Distractor
(3) Implausible Sentence, Plausible Distractor
(4) Implausible Sentence, Implausible Distractor

Figure 2. *Total Viewing Times in Milliseconds for Experiment 2*



(1) Plausible Sentence, Plausible Distractor
(2) Plausible Sentence, Implausible Distractor
(3) Implausible Sentence, Plausible Distractor
(4) Implausible Sentence, Implausible Distractor