Published in final edited form as:

*J Exp Psychol Learn Mem Cogn*. 2015 September ; 41(5): 1497–1515. doi:10.1037/xlm0000119.

# Subsequent to suppression: Downstream comprehension consequences of noun/verb ambiguity in natural reading

**Mallory C. Stites** and **Kara D. Federmeier**

University of Illinois, Urbana-Champaign

## Abstract

We used eye-tracking to investigate the downstream processing consequences of encountering noun/verb (NV) homographs (i.e., *park*) in semantically neutral but syntactically constraining contexts. Target words were followed by a prepositional phrase containing a noun that was plausible for only one meaning of the homograph. Replicating previous work, we found increased first fixation durations on NV homographs compared to unambiguous words, which persisted into the next sentence region. At the downstream noun, we found plausibility effects following ambiguous words that were correlated with the size of a reader's first fixation effect, suggesting that this effect reflects the recruitment of processing resources necessary to suppress the homograph's context-inappropriate meaning. Using these same stimuli, Lee and Federmeier (2012) found a sustained frontal negativity to the NV homographs, and, on the downstream noun, found a plausibility effect that was also positively correlated with the size of a reader's ambiguity effect. Together, these findings suggest that when only syntactic constraints are available, meaning selection recruits inhibitory mechanisms that can be measured in both first fixation slowdown and ERP ambiguity effects.

### Keywords

eye-tracking; ERPs; sentence processing; noun/verb homographs; plausibility

## Introduction

Ambiguity is a core feature of language and, accordingly, there is now a large body of work, using multiple measures, that has looked at the processing consequences of lexical ambiguity when it is first encountered. Much less clear, however, are the downstream effects of ambiguity and the extent to which these are subject to context-based or individual differences. When are multiple meanings activated, and at what point (if ever) does one of them become suppressed? How effective is suppression when it is used? What representations are readers left with after they have seemingly disambiguated ambiguous words and moved on in the sentence? The literature looking at downstream effects after disambiguation is sparse (Gunter, Wagner & Friederici, 2003; Hagoort & Brown, 1994; Lee & Federmeier, 2012; Miyake, Just & Carpenter, 1994), and so these questions remain

Address correspondence to: Mallory C. Stites, Department of Psychology, University of Illinois, 603 E. Daniel St., Champaign, IL 61820, Phone: 636.208.5809, stites2@illinois.edu.

largely unanswered. In the present work, we use eye-tracking to conduct a first study of the downstream effects of lexical ambiguity resolution by young adults during natural reading, in order to better understand what happens after the initial ambiguity costs are observed.

The bulk of previous research on the processing of ambiguous words has focused on noun/noun (NN) homographs, whose meanings fall within the same syntactic class (e.g., *calf*). One of the most robust findings in the literature is that encountering lexical ambiguity often creates processing burdens in the form of increased reading time for ambiguous words relative to matched controls. Empirical work and theorizing has focused on uncovering the nature of these costs and factors that can ameliorate them. For example, the Reordered Access Model (Duffy, Morris, & Rayner, 1988) proposes that readers exhaustively access all of an ambiguous word's meanings upon first encountering it, but that factors like the preceding context and the frequency of the each sense can affect the order with which these meanings become active. Thus, a subordinate-biasing context will increase the activation given to a biased homograph's less frequent meaning, causing it to compete for selection with the dominant meaning and increasing reading times on the word (the *subordinate bias effect*, Rayner, Pacht, & Duffy, 1994). Reading times are also longer on balanced homographs with two equally frequent meanings, relative to unambiguous words, when preceded by neutral contexts (Rayner & Frazier, 1989), suggesting that the two meanings compete for selection in the absence of a biasing context. An extensive body of research has explored these individual and joint effects of meaning frequency and context on the processing of NN homographs (for review, see Duffy, Kambe, & Rayner, 2001).

Of particular interest to the current study is how readers resolve the cross-class ambiguity associated with noun/verb (NV) homographs, a single written wordform that has both a noun and a verb sense. A line of ERP work has used NV homographs to examine the processing consequences of cross-class ambiguity and to probe how not only semantic but also syntactic cues are each used to resolve that ambiguity. Specifically, several studies by Federmeier and colleagues (Federmeier, Segal, Lombrozo, & Kutas, 2000; Lee & Federmeier, 2006, 2009, 2012) have found that NV homographs elicit a sustained negativity over frontal channels (relative to unambiguous words) when they appear in semantically neutral but syntactically constraining contexts. Importantly, this negativity is only present for NV homographs with two semantically distinct meanings (i.e., *to/the park*), but not when the two senses overlap in meaning (i.e., *to/the vote*) (Lee & Federmeier, 2006). When semantic constraints are available, NV homographs are processed in a qualitatively similar manner as unambiguous words -- that is, without eliciting the frontal negativity. Even in semantically unambiguous contexts, though, the dominant meaning of the ambiguous word remains somewhat active, as evidenced by larger N400 responses to subordinate (but not dominant) instantiations relative to unambiguous controls (Lee & Federmeier, 2009).

Given the prevalence of ambiguity effects for NV homographs in ERP work, it is surprising that previous eye-tracking studies have failed to find comparable reading time costs (Folk & Morris, 2003). This has led to a theoretical discrepancy across the literatures as to whether syntactic cues alone are insufficient to constrain lexical access (as suggested by the ERP work) or can allow readers to selectively activate only the context-appropriate meaning (as suggested by the eyetracking work). To address this discrepancy, Stites, Federmeier, and

Stine-Morrow (2013) measured eyegaze during natural reading using the stimuli from a previous ERP study (Lee & Federmeier, 2009). They embedded NV homographs at the end of two context types: congruent sentences, which have both semantic and syntactic cues to the context-appropriate meaning (e.g., You can usually find the registration desk of a hotel in the **lobby**), and syntactic prose sentences, which maintain the syntactic cues of congruent sentences but lack coherent semantics (e.g., You can usually install the math student of a day in the **lobby**). Eye-tracking results showed increased first fixation durations on the NV homographs in the contexts with constraining syntax but lacking coherent semantics (the same contexts that elicited the frontal negativity), but did *not* find similarly increased reading times in the semantically rich contexts.

Interestingly, the first fixation and frontal negativity ambiguity effects share several other features beyond their eliciting contexts. First, the timing of their onsets is similar, both coming online around 200-ms after readers first apprehend the target word. Secondly, older adults as a group fail to elicit both the frontal negativity (Lee & Federmeier, 2011) and the first fixation effects. Rather, older adults spend more time *rereading* the NV homographs after leaving them in the first pass, suggesting incomplete initial ambiguity resolution. Finally, both effects show a similar correlation with individual verbal fluency scores. Lee and Federmeier (2011) found a positive correlation between participants' verbal fluency and the size of their frontal negativity effect, and a median split found that older adults in the high verbal fluency group exhibited a young-like pattern of effects. In Stites et al. (2013), both young and older adults with higher verbal fluency showed larger first fixation effects and smaller rereading effects, suggesting a tradeoff between early and late disambiguation that was related to their verbal fluency. Because verbal fluency tests are believed to at least partially reflect frontal lobe efficacy (Henry & Crawford, 2004; Stuss & Levine, 2002), the evidence so far suggests that the first fixation and frontal negativity effects both reflect the recruitment of frontally-mediated meaning selection mechanisms needed to disambiguate NV homographs in the absence of constraining semantics (cf. Novick, Trueswell, & Thompson-Schill, 2010, and Thompson-Schill, Bedny, & Goldberg, 2005). Taken together, then, the behavioral (Rodd, Gaskell, & Marslen-Wilson, 2002), ERP (Federmeier et al., 2000; Lee & Federmeier, 2006, 2009, 2012), and eye-tracking (Stites et al., 2013) data provide strong support for the idea that in the absence of semantic constraints, additional processes must be recruited to aid ambiguity resolution.

Although the effects of ambiguity at the point it is first encountered have now been fairly extensively studied and incorporated into theoretical accounts of ambiguity resolution, much less clear are the consequences of ambiguity for downstream processing. Given that the target words used in the stimuli for Lee and Federmeier (2009) and Stites et al. (2013) appeared in the sentence-final position, we do not have evidence as to the activation level of the context-inappropriate word at the point *after* the ambiguity effects were elicited. As such, we cannot definitively rule out the possibility that the ambiguity effects observed may have reflected *maintenance* of both meanings in working memory instead of *selection* of just the context-appropriate meaning. In fact, previous ERP work has proposed a link between certain sustained frontal effects and working memory processing during language comprehension, specifically in the construction of mental models (Kutas & King, 1996) and

the reanalysis required for joke comprehension (Coulson & Kutas, 2001). In the eye-tracking domain, Filik and Moxey (2010) found that longer reading times on ambiguously ironic versus literal sentence regions resulted in the continued activation of both the literal and ironic meaning downstream--suggesting that the reading time increase reflected active maintenance of both meanings in working memory. As such, it is important to establish the functional significance of the ambiguity effects observed by Lee and Federmeier (2009) and Stites et al. (2013) with respect to the consequences they have for downstream processing, to add additional evidence to our claim that they reflect meaning *selection* rather than *maintenance*.

Lee and Federmeier (2012) examined this question by presenting NV homographs in sentences that were preceded by a semantically neutral but syntactically constraining phrase (i.e., *Alan wanted to/the train*...). The target word was followed by a prepositional phrase, containing a noun (shown here in bold) that was plausible for only one of the homograph's senses. For example, the sentence Alan wanted to train at the **gym** every day until he lost his excess weight is plausible, whereas the sentence Alan wanted the train at the **gym** every day until he lost his excess weight is not. Thus, the downstream activation of the context-appropriate word can be probed naturally within the sentence context through a plausibility manipulation. This stands in contrast to the more traditional probe technique in which meaning activation is measured via the presentation of a word that is semantically related to one of the homograph's meanings, which Van Petten and Kutas (1987) found may actually produce backward priming of the related meaning rather than simply testing its current activation level. Lee and Federmeier (2012) hypothesized that if the frontal negativity reflects meaning selection, then N400 plausibility effects (i.e. reduced N400 amplitudes for plausible relative to implausible words) should be present at the downstream noun following both ambiguous and unambiguous words, indicating downstream activation of only the context-appropriate meaning. On the other hand, equal facilitation for the plausible and implausible nouns following ambiguous words would indicate that both meanings are still active, suggesting that the frontal negativity reflects meaning maintenance rather than selection.

Lee and Federmeier (2012) replicated the sustained negativity previously found to NV homographs, beginning around 200-ms after target word onset and continuing through the first two words of the following prepositional phrase (henceforth, the preposition region). Importantly, at the downstream noun, they found an N400 plausibility effect following both ambiguous and unambiguous words. For young adults, this effect was numerically smaller in nondominant contexts, suggesting residual activation of the dominant-related meaning, despite the unambiguous syntactic cues. Older adults as a group did not elicit the frontal negativity effect, and at the downstream noun they showed continued activation of the dominant meaning in both dominant- and nondominant-biased contexts. Furthermore, Lee and Federmeier (2012) also found that for both young and older adults, the size of one's frontal negativity effect at the homograph was positively correlated with the size of their plausibility effect at the later noun. This pattern suggests that the frontal negativity reflects suppression of the homograph's context-inappropriate meaning, making it less available

downstream and increasing the size of the reader's N400 effect to words plausibly related to it.

The ERP pattern offers strong evidence that the frontal negativity effect reflects meaning selection, rather than meaning maintenance, of NV homographs in the absence of constraining semantics. By using eye-tracking during natural reading, the current study will extend these results and build on the evidence that the first fixation effect is sensitive to the same underlying processes as the frontal negativity. The ERP evidence shows that the young adult brain can effectively suppress context-irrelevant meanings of words under conditions wherein words are presented serially and at a relatively slow pace. The use of eye-tracking affords the benefit of observing how readers behave when they determine what to look at, set the pace at which they take in information, and have the opportunity to look back in the text. For example, readers tend to skip function words as much as 67% of the time (Rayner, 1998). Although word skipping is often driven by oculomotor effects -- shorter words are skipped more frequently than longer words (Rayner, Slattery, Drieghe, & Liversedge, 2011) – it can also be driven by contextual effects, as more predictable words are skipped more often than less predictable words (Fitzsimmons & Drieghe, 2013; Rayner, et al., 2011). Moreover, several studies have found that readers are less likely to skip a parafoveal nonword than a real word, even when it is orthographically very similar to the expected word (Drieghe, Rayner, & Pollatsek, 2005; Gordon, Plummer, & Choi, 2013). It is typically assumed, then, that readers tend to skip words they have already comprehended before actually fixating on them. It is interesting, then, that the ambiguity effects observed in Stites et al. (2013) were most pronounced when readers *skipped* the function word preceding the homograph (which was the only cue to the word's intended meaning), indicating their need to recruit the frontally-mediated meaning selection mechanisms to correctly interpret the word. If readers skipped the function words because they had *already* comprehended them, it is even more striking that they are unable to then *use* that information in advance to selectively activate the homograph's context-appropriate sense.

In the current materials, as in Stites et al. (2013), a function word provides the only advance information about a homograph's context-appropriate meaning. Will reader's natural tendency to skip these short -- often predictable -- words diminish his or her ability to disambiguate the homographs? Assuming readers do take in the syntactic information, do they take the extra time required to suppress the context-inappropriate meaning of ambiguous words, and when do they first initiate suppression? Finally, how will individuals' tendency to suppress or not affect their later processing of the downstream noun, including their subsequent rereading of earlier parts of the sentence?

We will answer these questions by using the stimuli from Lee and Federmeier (2012), giving us the opportunity to compare natural reading patterns with ERP effects previously elicited by the same contexts to look for correspondences and differences between measures. We predict to again see increased first fixation durations to the NV homographs relative to unambiguous words, conceptually replicating Stites et al. (2013). Given that we see the first fixation effect, we will test whether not only the fixations on the target homograph, but also those continuing into the preposition region will prove to be susceptible to slowdown by the processes reflected in the frontal negativity, which was found to be sustained into this

region. Critically, the current set of materials will also allow us to examine, for the first time, the downstream consequences of ambiguity resolution during natural reading. In particular, we will examine the nature of the plausibility effect at the noun of the prepositional phrase and its relationship (if any) to effects observed at the prior NV homograph itself.

Previous eye-tracking studies investigating plausibility effects have used a wide range of plausibility manipulations, from violating selectional or animacy restrictions of verbs (Warren & McConnell, 2007) to pairing unlikely instrument-verb-object triplets (Patson, & Warren, 2011; Matsuki, Chow, Hare, Elman, Scheepers, & McRae, 2011; Rayner, Warren, Juhasz, & Liversedge, 2004), to invoking a fantasy and/or fictional context for literally anomalous words (Filik, 2008; Warren, McConnell, & Rayner, 2008). Perhaps as a consequence, the literature does not provide a clear answer about when and how plausibility effects first come on-line in eye-tracking measures. Some studies have reported effects of *plausibility* (not possibility) on first fixation and gaze durations on the violating word itself (Matsuki et al., 2011; Patson & Warren, 2011 [Experiment 2]; Warren, McConnell, & Rayner, 2008), whereas others have found that plausibility effects emerge only on later reading measures on the violating word or fixations on the next region in the sentence (Filik, 2008; Patson & Warren, 2011 [Experiment 1]; Rayner, et al., 2004; Warren & McConnell, 2007). It has been posited that the degree of the plausibility violation may determine the timecourse with which it first affects the eye movement record, with more severe or impossible violations eliciting earlier effects (i.e., first fixation or gaze duration to the violating word), and less severe plausibility violations causing later disruptions (i.e., go-past time, regressions out, spillover) (Warren, 2011). It has also been shown that both the context in which the implausibility appears, as well as the featural overlap between the plausible and implausible/impossible word can sometimes mitigate early or late effects of implausibility (see Warren, 2011, for a discussion). However, the severity of a given manipulation is not well-defined across previous experiments, and the predictability of the plausible versus implausible/impossible words is rarely normed, making it difficult to assess the locus of effects across diverse sets of materials. Attempts to model the exact conditions that will produce longer reading times versus regressions have produced variable results (Reichle, Warren, & McConnell, 2009), making it difficult to predict the specific pattern of eye movement effects that will be elicited by the current materials.

In contrast, ERP studies that manipulate plausibility find robust N400 plausibility effects elicited by the violating word itself, regardless of the type of violation used (Ferguson, Sanford, & Leuthold, 2008; Filik & Leuthold, 2008; Lee & Federmeier, 2012; Nieuwland, 2013; Nieuwland & Kuperberg, 2008; Nieuwland & Martin, 2012; Nieuwland & Van Berkum, 2006; Paczynski & Kuperberg, 2012) -- but in a timeframe (200-500 ms post-stimulus-onset) that, in natural reading, would span early gaze measures on the violating word and later gaze measures/fixations onto subsequent words. For the purposes of the current study, the *pattern* of plausibility effects, rather than the timing of their onset, is of import. Therefore, we will combine the violating word and its subsequent spillover word into a single region, allowing us to capture the widest range of plausibility effects, including slowing down on and/or regressing back in the text from either one of these words.

Of critical interest for the present study is whether the plausibility effects observed downstream will differ following ambiguous and unambiguous words. Based on our evidence that the first fixation effect, like the frontal negativity effect, reflects meaning selection, we predict that we will see plausibility effects of roughly equal magnitude following ambiguous and unambiguous words, as Lee and Federmeier (2012) observed in their ERP study. Importantly, to the extent that the first fixation and frontal negativity effects both reflect meaning selection of the NV homograph, we also predict that we will see a relationship between the size of the first fixation and plausibility reading time effects for sentences containing ambiguous words. If we find a positive correlation between these effects in eye-tracking, as was observed in the ERP data, we can feel confident that these two effects likely reflect a similar underlying source and functional role, namely the recruitment of frontally-mediated selection mechanisms to aid in difficult ambiguity resolution.

## Methods

### Participants

Twenty-four University of Illinois undergraduate students (19 females and 5 males) with a mean age of 19 years (range: 18-21) participated in the study for course credit. All were native English speakers and were right-handed as assessed by the Edinburgh inventory (Oldfield, 1971)

### Materials

The materials and design employed were identical to that in Lee and Federmeier (2012) to allow for the closest comparison between the two studies. Two word types were used as targets: noun/verb (NV) homographs, which have both a noun and a verb meaning (e.g., *trip*), and matched unambiguous words (e.g., *lady*). The two sets of target words were matched on log frequency (Kucera & Francis, 1967), length, and usage-specific concreteness, which are listed in Table 1. Homographs were also rated on semantic distinctiveness (the similarity the noun and verb senses of the homograph), as judged by participants in a previously conducted norming study (Lee & Federmeier, 2006). Only homographs with high semantic distinctiveness (i.e., dissimilar noun and verb senses) were used.

In order to compare effects amongst the NV homographs based on meaning dominance, a dominance score was calculated for each ambiguous word. We did so by subtracting the log frequency of its noun usage from its verb usage, and dividing this number by the total frequency count, with usage-specific frequency data collected from WordNet 3.0 (Princeton University, 2010; <http://wordnet.princeton.edu>). This produced a value ranging from −1 to 1 (with −1 being totally noun-biased, and 1 being totally verb-biased). For a given item, if the dominance score was greater than |.3|, it was considered biased; if its bias score fell between −.3 and .3, it was considered balanced. Using this procedure, 33 of the 48 NV homographs were classified as biased, whereas the other 15 were balanced. The average dominance score for the biased words was |.69|, whereas the balanced words scored an

average of |.13|. For 20 of the 33 biased words, the noun sense was the dominant meaning, whereas for the other 13 biased words, the verb sense was dominant.

Target words were presented in sentence-medial position preceded by a semantically neutral but syntactically constraining phrase (i.e., *John hated to/the trip...*). They were followed by a prepositional phrase, which contained a noun that was either plausible or implausible given the target word (downstream noun frequency [Kucera & Francis, 1967]: *M*=91.5, range=0-909; length: *M*=5.7, range=3-11). For sentences containing NV homographs, the downstream noun could be related to either meaning of the homograph. For example, for the target word *rat*, one sentence was created instantiating the noun meaning (James hated the rat on the **floor** and tried to chase it around the room), and one which instantiated the verb meaning (*Roy hated to rat on his **sister** for stealing*). The sentences were made implausible by swapping the syntactic cue preceding the homograph (i.e. *to* and *the*) while leaving the rest of the sentence identical (e.g., *James hated* to *rat on the* **floor** *and tried to chase it around the room*; *Roy hated* the *rat on his* **sister** *for stealing*) (see Table 2, or the Appendix for more example stimuli). This change rendered the downstream noun (*floor*, *sister*) implausible with the context-appropriate meaning, but still plausible with the context-*inappropriate* meaning of the target word. In this way, the same set of nouns served as both plausible and implausible continuations following the ambiguous words. Similar plausible and implausible sentences were created for the unambiguous target words as well, by replacing the downstream plausible noun with an implausible one (e.g., *Gary planned to carve out the {**pumpkin / meeting**} and put it on his front stoop*). These two sets of downstream nouns did not differ from each other in terms of length (plausible: *M*=6.3, range=3-13; implausible: *M*=6.2, range=3-13; $t(95)=-0.45$, $p=.65$) or frequency [Kucera & Francis, 1967] (plausible: *M*=106, range=0-1207; implausible: *M*=116, range=0-1207; $t(95)=0.40$, $p=.69$). All stimuli were controlled so that there was little semantic association between the target word and the downstream noun (according to the University of South Florida free association norms, both forward and backward association strength was 0.01), allowing us to avoid semantic priming effects on the downstream noun.

The sentential constraint up to the downstream noun, as well as the cloze probability of the noun itself, did not differ across conditions. Sentential constraint was defined as the cloze probability of the most popular sentence completion supplied in a paper-and-pencil norming study (originally reported in Lee & Federmeier, 2012), whereas cloze probability refers to the percentage of respondents who supplied that particular item as the best completion. The sentential constraint leading up to the head nouns was matched following unambiguous words (plausible: *M*=24%, *SD*=14; implausible: *M*=24%, *SD*=14) and NV homographs (plausible: *M*=25%, *SD*=15; implausible: *M*=25%, *SD*=15). Further, the cloze probability of the downstream nouns did not differ across conditions (following unambiguous words, plausible: *M*=12%, *SD*=15; implausible: *M*=0%, *SD*=0 and following NV homographs, plausible: *M*=16%, *SD*=17; implausible: *M*=0%, *SD*=0). There were no significant differences between these measures across conditions (all *p*'s >.1; Lee & Federmeier, 2012). To ensure the validity of the plausibility manipulation, the plausibility of the downstream nouns used in the study was rated as over 75% for the plausible sentences (unambiguous sentences: *M*=80%, *SD*=14; ambiguous sentences: *M*=76%, *SD*=20) and less than 35% for

the implausible sentences (unambiguous sentences: *M*=29%, *SD*=16; ambiguous sentences: *M*=33%, *SD*=17) (see Lee & Federmeier, 2012 for further details about the norming procedures used to create the stimuli).

Participants read each ambiguous and unambiguous word in two different sentences, with a lag of at least 50 sentences between them, and always followed with different continuations. To make the plausibility of the sentence unpredictable, across lists the two appearances of the word were either both plausible, both implausible, or one of each. The order of each pair of sentences was also counterbalanced, creating a total of eight lists. Each participant read 192 sentences, with equal numbers of sentences containing ambiguous and unambiguous target words, each with the same number of plausible and implausible continuations. Additionally, ambiguous words were used equally in their noun and verb sense, and the same number of unambiguous nouns and verbs appeared as well (although the factor of grammatical class will not be examined in this study).

## Procedure

Participants were tested individually in a quiet room, seated 71 cm away from a 22" Cornerstone p1750 CRT monitor. Eye movements were recorded using an SR Research Eyelink 1000 desk-mounted eye-tracker with a sampling rate of 1000 Hz. A chinrest was used to encourage participants to minimize head movements during the experiment. To ensure accurate tracking, a nine-point calibration procedure was used at the start of the experiment, and after every break or movement out of the chinrest. Additionally, a drift-check was performed at the start of every trial, and participants were recalibrated if the drift-check failed. Recordings were monocular, taken from the right eye.

At the start of the experiment, participants were presented with written instructions, after which they received seven practice trials to acclimate them with the experimental procedure. Each trial began with the presentation of a drift check target in the upper left corner of the screen. Participants fixated the target while pressing the "start" button on a hand-held controller. At this time the sentence appeared, which was presented in left-justified, black Courier New font on a light gray background. Three characters subtended one degree of visual angle. Participants were instructed to read the sentence normally, and press the "advance" button on the controller when they were finished reading the sentence. After each sentence, a probe word appeared in uppercase red font in the center of the screen. Participants indicated via button-press if the word was presented in the immediately preceding sentence or not. Half of the probe words were new; the old probe words were chosen from syntactic cues, NV-homographs and matched unambiguous words, critical nouns, and the rest of the content words in the sentence.

The experiment was divided into four blocks of 48 trials, each of which took approximately 8 minutes to complete. At the end of each block, participants were allowed to remove their heads from the chinrest. A paper-and-pencil sentence recognition test was also administered at this time. The sentence recognition test consisted of 96 sentences total, half of which were new; the old sentences were drawn equally from plausible and implausible sentences containing NV homographs and matched unambiguous words. Participants were told to mark which sentences were present in the immediately preceding block. Because many of

the same words were used across the new and old sentences, simple word recognition would not allow good performance on this test. The word and sentence recognition tasks were included to keep task demands as similar as possible to Lee and Federmeier (2012), as well as to encourage participants to attend closely to every sentence presented, even implausible ones.

### Data Analysis

Within the Eyelink 1000 data analysis package, consecutive fixations less than 80 ms in duration and less than .5° apart in visual angle were combined into one fixation. Single fixations that were shorter than 80 ms or longer than 800 ms were then automatically excluded. Fixations shorter than 80 ms are unlikely to represent meaningful cognitive processing (Rayner, 1998), and fixations longer than 800 ms are likely the result of cases in which the tracker temporarily lost the eye, causing inaccurately inflated reading times. After this automatic exclusion was performed, an additional 1.1% of trials were excluded from the analysis due to track loss or program error. The percentage of trials lost on each interest area due to failure of the reader to fixate the region in the first pass is as follows: target word—3.1%, preposition region—0.6%, head noun region—0.3%. Additionally, the percentage of trials on which these regions were never fixated is as follows: target word—9.5%, preposition region—8.1%, head noun region—2.4%.

## Results

### Behavior

The purpose of the word and sentence recognition tasks was to encourage participants to pay close attention to the experiment at both the word and sentence level. Overall, participants answered the word recognition task with 91% accuracy (d'=2.8). Performance on the sentence recognition task was also good, at 82% (d'=2.2). Both of these values are comparable to how the young adults in Lee and Federmeier (2012) scored on these same tasks: average accuracy of 94% (d'=3.4) on the word recognition task and 86% (d'=2.5) on the sentence recognition task.

### Eye-tracking

Based on the findings of Lee and Federmeier (2012), several areas of interest were examined in the current dataset (see Figure 1 for a graphical depiction of the interest areas). First, we examined reading times on the target ambiguous word, to determine if there were ambiguity effects for the NV homographs relative to the unambiguous words, particularly on first fixation duration as observed by Stites et al. (2013). Next, we considered reading times on the first two words of the prepositional phrase following the targets (i.e., the preposition region), to look for additional first fixation ambiguity effects in this region as a corollary to the sustained negativity observed to this region by Lee and Federmeier (2012). Reading times on the downstream noun region (including the noun in the prepositional phrase and the word immediately following it) were then examined to test for a plausibility effect (i.e. longer reading times) on the implausible relative to plausible words, as well as an interaction between plausibility effects and preceding target word ambiguity.

Several standard eye-tracking measures were employed: first fixation duration, the length of the readers' first fixation on the word/region; gaze duration (a.k.a. first-pass time), the sum of all fixations on a word/region the first time it is fixated, before the eyes leave it in either direction; go-past time (a.k.a. regression path duration), a right-bounded measure including all fixations a reader makes on a word/region beginning with their first fixation until they move past it to the right (including rereading earlier parts of the sentence); and rereading time, all time spent reading a word/region after leaving it in the first pass. For all regions of interest, two repeated measures Analyses of Variance (ANOVA) were conducted, one averaging over participants ($F_1$) and the other averaging over items ($F_2$). For the by-participants analysis, both the ambiguity (or dominance) of the target word and the plausibility of the downstream noun (when applicable) were within-participant factors. For the by-items analysis, the identity of the NV homograph and/or matched unambiguous control was considered as the item over which to average (considering as separate lexical items the dominant and subordinate instantiations of the biased homographs, and the noun/verb meaning of the balanced homographs), making ambiguity (or dominance) a between-items factor. Because all lexical items were used both plausibly and implausibly, the factor of plausibility is a within-items factor.

We report first-pass reading time measures (first fixation, gaze duration, and go-past time) separately from rereading time measures. This allows us to present the results in a manner that follows naturally from the ERP study, because readers only receive a "first pass" through the sentence when using the RSVP presentation style used in ERP research. Furthermore, the factor of plausibility does not become apparent until readers reach the downstream noun region. As such, plausibility must be included as a factor in analyses of rereading times, but not first pass measures, of the target region.

## First-Pass Reading Measures

**Target Word:** Reading times on the target word (listed in Table 3) were analyzed using one-way ANOVAs with the factor of ambiguity (ambiguous vs. unambiguous), which was a within-participant factor for F1 analyses and a between-items factor for F2 analyses. First fixation durations showed a significant ambiguity effect, $F_1(1,23)=6.52$, $p<.05$, $F_2(1,94)=5.23$, $p<.05$, with longer first fixations for ambiguous relative to unambiguous words (see Figure 2). As in Stites et al. (2013), this effect was driven by cases in which participants landed directly on the target word without first fixating the preceding function word (which occurred on 55% of ambiguous trials and 57% of unambiguous trials), $F_1(1,23)=10.64$, $p<.01$, $F_2(1,94)=9.61$, $p<.01$. When participants landed on the function word first, there was no ambiguity effect on their subsequent first fixation to the target word, $F_1$ and $F_2$ <1. This is likely because the function word was short enough to allow readers to process both the function and target word simultaneously while fixating it, giving readers the opportunity to disambiguate the target word on their previous fixation. Correspondingly, there was no ambiguity effect on target word gaze duration, $F_1(1,23)=1.76$, $p=.20$, $F_2<1$. These findings replicate Stites et al. (2013), who found longer first fixation durations (but not gaze durations) on NV homographs compared to unambiguous words in syntactically constraining contexts that lacked coherent semantics, which were the same contexts that also elicited the sustained frontal negativity to the NV homographs (Lee and Federmeier, 2009,

2012). It is important to note that most targets received only one first-pass fixation; first-pass refixations happened on 15% of ambiguous targets and 20% of unambiguous targets. However, there was an ambiguity effect for go-past time, $F_1(1,23)=8.65$, $p<.01$, that did not reach significance by items, $F_2(1,94)=1.52$, $p=.22$, indicating that readers spent more time rereading earlier parts of the sentence before moving past ambiguous than unambiguous words.

In addition to overall ambiguity effects, we further explored the first fixation effect as a function of the target word's meaning dominance using one-way ANOVAs with four levels of dominance (dominant, subordinate, balanced, unambiguous), which was a within-participants but between-items factor. First fixations on the subordinate instantiations of the NV homographs ($M$=229, $SD$=34) were longer than the other three conditions, which were very similar to one another (dominant: $M$=218, $SD$=32; balanced: $M$=220, $SD$=31; unambiguous: $M$=216, $SD$=25). Omnibus ANOVAs revealed a main effect of dominance, $F_1(3,69)=5.53$, $p<.01$, $F_2(3,140)=3.46$, $p<.05$, indicating an overall difference between conditions. Follow-up tests were conducted to determine whether the first fixations were especially inflated for the subordinate instantiations, which would be the first piece of evidence for a subordinate bias effect for NV homographs. A paired samples t-test revealed that first fixations were longer to the subordinate instantiations of ambiguous words than to unambiguous words, both by participants, $t_1(23)=4.03$, $p<.01$, and by items (95% CI $-11.5 \pm 9.4$); the similar effect for subordinate compared to dominant instantiations was significant by participants, $t_1(23)=3.08$, $p<.01$, but failed to reach significance by items (95% CI $8.4 \pm 10.3$). Neither dominant instantiations nor balanced homographs differed from the unambiguous words (dominant: $t_1(23)=.053$, $p=.61$, by-items (95% CI $3.15 \pm 9.4$); balanced: $t_1(23)=1.12$, $p=.27$, by-items (95% CI $3.52 \pm 9.7$). These results are thus suggestive that the first fixation effect was more pronounced when the sentence invoked the subordinate meaning of the homograph.

**Preposition Region:** To test for the presence of ambiguity effects in the post-target region, as seen in the ERP study, we next examined reading times on the preposition region immediately following the target word. Because this region consisted of two short function words, which have an estimated skipping rate of 65% in normal reading (Rayner, 1998), the two-word phrase was considered as a single region for all analyses. Again, one-way ANOVAs were constructed with the factor of ambiguity of the preceding target word, which was manipulated within-participants but between-items. Results showed a significant ambiguity effect on the first fixation to the preposition region, $F_1(1,23)=5.74$, $p<.05$, $F_2(1,94)=4.18$, $p<.05$, with longer reading times following ambiguous than unambiguous words (see Figure 2). There was no ambiguity effect for gaze duration on the region, $F_1$ and $F_2$ <1, although there was again a significant ambiguity effect for go-past time by participants, $F_1(1,23)=7.33$, $p<.05$, that did not reach significance by items, $F_2(1,94)=2.01$, $p=.16$. The presence of the ambiguity effect on first fixations in the preposition region indicates that this measure is particularly sensitive to the processes reflected in the frontal negativity effect (which was sustained through this region), such that the processes generating the negativity also work to slow first fixations. Given that gaze durations on the preposition region itself were not longer following ambiguous words, and considering that

go-past time includes reading times on this region *and* everything to its left before readers move forward in the text, the increased go-past times on the region reflect time spent rereading earlier parts of the sentence before moving forward, possibly on the NV homograph or its preceding function word.

**Downstream Noun Region:** Next, first-pass reading times were examined for the downstream noun in the prepositional phrase to test for plausibility effects. Recall that the implausible words following the NV homographs were in fact plausible with the homograph's context-*inappropriate* meaning, so longer reading times in the implausible than plausible condition would imply successful suppression of the alternate meaning. If, on the other hand, the first fixation effect indexes active maintenance of the context-inappropriate meaning of the NV homograph, we would expect reading times on plausible and implausible words to be similarly facilitated following the homographs, indicating continued activation of the alternate meaning. There were two factors of interest in this region: the ambiguity of the preceding target word and plausibility of the downstream noun itself. Accordingly, 2×2 repeated measures ANOVAs were constructed: for F1 analyses, both ambiguity (ambiguous vs. unambiguous) and plausibility (plausible vs. implausible) were within-participant factors, whereas in the F2 analyses, plausibility was manipulated within-items but ambiguity was still a between-items factor. Overall, first-pass reading times on the noun region showed significantly longer reading times on implausible relative to plausible nouns, with no consistent effects of preceding ambiguity or interaction between the two factors (see Table 3, and Figure 3). First fixations to the downstream noun region exhibited a significant effect of plausibility, $F_1(1,23)=11.40$, $p<.01$, $F_2(1,94)= 10.85$, $p<.01$, with no effect of ambiguity, $F_1(1,23)=1.35$, $p=.26$, $F_2<1$, or interaction, $F_1$ and $F_2$ <1. Gaze durations again showed a main effect of plausibility, $F_1(1,23)=25.02$, $p<.01$, $F_2(1,94)=13.95$, $p<.01$, as well as a main effect of ambiguity that was significant by participants, $F_1(1,23)=13.94$, $p<.01$, but not items $F_2(1,94)=1.52$, $p=0.22$, and no interaction, $F_1$ and $F_2$ <1. Finally, go-past times showed a similar pattern, in which only the effect of plausibility was significant, $F_1(1,23)=26.37$, $p<.01$, $F_2(1,94)=25.78$, $p<.01$, with no effect of ambiguity, $F_1(1,23)=1.60$, $p=.22$, $F_2<1$, nor interaction, $F_1(1,23)=1.807$, $p=.19$, $F_2(1,94)=1.29$, $p=.26$.

Readers appeared to quickly appreciate the (im)plausibility of the downstream noun as reflected by increased first pass reading times on these words, and this initial plausibility assessment was unaffected by whether the noun was preceded by an ambiguous word. This is an important point to note, as the implausibility of the noun following ambiguous words was only apparent if readers sufficiently suppressed the context-inappropriate meaning of the NV homograph. The observed pattern of results mirrors the findings from Lee and Federmeier (2012), in which young adults showed N400 facilitation for plausible relative to implausible nouns following both ambiguous and unambiguous words. The lack of an interaction between ambiguity and plausibility across both studies indicates that readers were able to use syntactic cues alone to suppress the context-inappropriate meaning of the NV homograph well enough to recognize the implausibility of the downstream nouns in the ambiguous context.

## Correlations between ambiguity effects and plausibility effects

Lee and Federmeier (2012) found in their ERP study that the size of a reader's frontal negativity effect was positively correlated with the size of their downstream plausibility effect. This correlation provided evidence that the frontal negativity reflected meaning selection (rather than maintenance), because a larger initial ambiguity effect made the context-inappropriate meaning of the ambiguous word less available later in the sentence. We thus predicted that if the first fixation ambiguity effect reflects the same underlying processes as the frontal negativity, then the size of a reader's first fixation ambiguity effect would similarly predict the availability of the word's context-inappropriate meaning later in the sentence.

In order to test this prediction, we first calculated an ambiguity effect for each participant by subtracting average first fixation durations to the unambiguous targets from average first fixation durations to the ambiguous targets (Ambiguous – Unambiguous), producing a positive value for longer ambiguous reading times. We then calculated three separate plausibility effects for each participant, one for each of the three first-pass reading measures on the noun region following ambiguous words only. Each of these plausibility effects was created by subtracting reading times on the plausible from the implausible noun regions (Implausible – Plausible), again producing a positive number for longer implausible reading times. This procedure was carried out separately for first fixation, gaze duration and go-past times. To test for a relationship between the early and downstream effects, we calculated the correlation between the first fixation ambiguity effect and each of the three downstream plausibility effects, looking for positive correlations ($t > 1.72$). Results showed a significant correlation between the first fixation ambiguity effect and the go-past time plausibility effect for the noun region ($r$=.39, one-tailed $p$<.05), as can be seen in Figure 4. This correlation is even stronger ($r$=.48) if we use only the subset of first fixations in which readers landed directly on the target -- i.e., those showing the clearest ambiguity effects. There was no significant relationship between the ambiguity effect and the other two reading time measures on the noun region (first fixation: $r$=.06; gaze duration: $r$=-.13; $p$'s > .25). Because go-past time includes all reading of the noun region as well as *rereading* of earlier parts of the sentence before moving forward, it captures multiple patterns of plausibility effects, including (but not limited to): slowing down on the implausible word itself, slowing down on the spillover word, and launching a regression from either the downstream noun or spillover word. These different combinations of reading time and regression-based plausibility effects are missed in first fixation or gaze duration alone.

Importantly, we wanted to ensure that this correlation was specific to sentences containing ambiguous words, rather than reflecting a general tendency of certain readers to slow down for all difficulties equally (i.e., for both ambiguous and implausible words across the board). To do so, we created a plausibility effect for go-past times on the noun region in the *unambiguous* contexts. We then calculated the correlation between this new plausibility effect and the first fixation ambiguity effect, and found no significant relationship between these measures ($r$=−.23; $p$ = .86). Because readers may elicit patterns of plausibility effects that differ systematically between ambiguous and unambiguous sentences, we also tested for the presence of positive correlations between the first fixation ambiguity effect with other

reading time measures on the downstream noun region for the unambiguous sentences, and again found no relationships in the predicted direction (first fixation: $r$=.09; gaze duration: $r$=−.36; probability of regression: $r$=−.09; $p$'s > .34).

Finally, to directly test that the size of the correlation between the first fixation ambiguity effect with the go-past plausibility effect following *ambiguous* words is different from its correlation with the same effect following *unambiguous* words, we calculated Steiger's Z (Meng, Rosenthal, & Rubin, 1992) for these two correlations. This test statistic compares the size of two non-independent correlations that share a common variable (i.e., the first fixation ambiguity effect), taking into account the size of the correlation between each of the two unique variables with the shared variable, as well as the correlation between the two unique variables themselves (i.e., go-past time plausibility effect following *ambiguous* vs. *unambiguous* words). The output of this test confirmed that these two correlations are statistically different from each other, $Z$=2.40, $df$=21, $p$<.05. Thus, we can conclude that the first fixation ambiguity effect is selectively correlated with go-past time plausibility effects following *ambiguous* words only, suggesting that the first fixation effect indexed a suppression process that left the context-inappropriate meaning less available downstream. This correlation mirrors that observed in the ERP study, and provides further evidence that the first fixation reading time effect reflects the recruitment of meaning selection mechanisms needed to disambiguate NV homographs in the absence of constraining semantics.

## Rereading Time Measures

**Target Word Region—**In Stites et al. (2013), we found that certain readers, specifically older adults as a group and young adults with low verbal fluency, were more likely to reread the NV homograph after leaving it in the first pass. These rereading patterns were taken as evidence of incomplete ambiguity resolution, resulting in the need to return to the NV homograph after having moved on to later parts of the sentence. In light of these results, we also examined rereading times for the target word in the current study as another way to assess ambiguity effects. Rereading times were considered for the target word *region*, which we defined to include both the target word and the class-disambiguating function word immediately preceding it (see Figure 1), as readers may choose to return to either word (or both) to gather more information about the homograph's intended sense. Ambiguity and plausibility were both included as factors to investigate whether downstream plausibility had a differential effect on rereading times to the target word region. We calculated rereading time by subtracting gaze duration from total time, excluding all first-pass reading times from the analyses. If readers did not return to the region after the first pass, that trial was assigned a rereading duration of zero in this analysis to jointly capture the *probability* of refixating the target region together with the *amount* of time spent rereading the region. Overall rereading times on the target region were longest for the ambiguous words in implausible contexts ($M$=158, $SD$=118), followed by ambiguous words used plausibly, ($M$=106, $SD$=67), unambiguous words used implausibly ($M$=96, $SD$=75), and finally unambiguous words used plausibly, ($M$=90, $SD$=73). These values were subjected to two 2×2 repeated-measures ANOVAs, again with both ambiguity (ambiguous vs. unambiguous) and plausibility (plausible vs. implausible) as within-participant factors for the F1 analysis, and with

plausibility as a between-items factor and ambiguity as a within-items factor for the F2 analysis. Results showed a main effect of ambiguity, $F_1(1,23)=38.12$, $p<.01$, $F_2(1,94)=27.55$, $p<.01$, reflecting the overall longer rereading times for ambiguous than unambiguous words, a main effect of plausibility, $F_1(1,23)= 10.49$, $p<.01$, $F_2(1,94)=13.74$, $p<.01$, reflecting overall longer rereading times for implausible than plausible words, as well as a significant interaction, $F_1(1,23)=6.12$, $p<.05$, $F_2(1,94)=8.13$, $p<.01$. Follow-up t-tests examining the plausibility effect within each ambiguity condition indicated that rereading times were significantly longer for implausible than plausible ambiguous regions, $t_1(24)=3.05$, $p<.01$, $t_2(47)=4.89$, $p<.01$, but did not differ for the unambiguous regions, $t_1(24)=1.04$, $p=.31$, $t_2(47)<1$. Thus, in the implausible sentences containing ambiguous words, readers recognize the NV homograph as the source of the implausibility (because the implausible noun would have been plausible with the *other* meaning of the homograph), and selectively return to the homograph.. However, the eventual implausibility of the sentence did not affect rereading times for the unambiguous target regions.

Based on the suggestive subordinate bias effect observed on first-pass reading measures, rereading times on the ambiguous target word region were also examined as a function of meaning dominance (see Table 4 for reading times). Specifically, we were interested in whether readers continued to experience difficulties with the subordinate instantiations of the NV homographs, or whether the increased reading times in the first pass allowed readers to fully select the subordinate meaning, leading to later rereading times that are indistinguishable from the dominant condition. An ANOVA with the factors of plausibility (2: implausible × plausible) and meaning dominance (3: dominant, subordinate, and balanced) showed again that rereading times were longer for targets in implausible than plausible contexts, $F_1(1,23)=10.07$, $p<.01$, $F_2(1, 93)=18.42$, $p<.01$. There was also a main effect of dominance condition that was significant by participants and marginal by items, $F_1(2,46)=3.70$, $p<.05$, $F_2(2, 93)=2.37$, $p=.10$, as well as an interaction between the two, $F_1(2,46)=3.79$, $p<.05$, $F_2(2,93)=3.09$, $p=.05$.

Follow-up t-tests were conducted to test for the significance of the plausibility effect within each of the three ambiguous word dominance conditions (because, as previously mentioned, there was no difference for the unambiguous words). As can be seen in Figure 5, rereading times were significantly longer for implausible than plausible targets in both the dominant and balanced conditions (dominant: $t_1(23)=4.55$, $p<.01$, $t_2(32)= 4.18$, $p<.01$; balanced: $t_1(23)=2.61$, $p<.05$, $t_2(29)= 3.49$, $p<.01$), whereas subordinate-plausible rereading times were just as long as the subordinate-implausible condition, $t_1(23)=0.98$, $p=.34$, $t_2(32)=0.46$, $p=.65$. Focusing just on the comparison between the dominant and subordinate conditions, we can see that there was no difference between their implausible rereading times, $t_1(23)=0.14$, $p=.89$. The striking difference was with respect to the plausible rereading times, which were 56 ms longer for the subordinate than dominant condition, $t_1(23)=2.89$, $p<.01$[1]. This pattern shows that even when the subordinate target was followed by a noun that was

[1]Although a similar t-test could not be conducted by-items (because it constitutes a between-items test), we instead calculated Tukey's HSD using the MSW from the by-items ANOVA (q=3.68, MSW=5695, n=140), which produced a cutoff value of 51 ms for a significant difference at the p=.05 level. We can see that the 2 ms difference between the subordinate implausible and dominant implausible conditions is not significant, whereas the 56 ms difference between the subordinate plausible and dominant plausible is significant.

plausible with the context-appropriate meaning of the NV homograph, readers still reread the target word region for just as long as they did when it was followed by an implausible word. It could be the case that readers were less confident in their disambiguation of the subordinate instantiation (perhaps due to residual activation of the dominant meaning), or may have marked the homograph as a place of uncertainty, either of which would have encouraged readers to selectively return to the homographs in the subordinate-plausible condition.

**Downstream Noun Region:** We also considered rereading times on the downstream noun region, as a way to measure continued disruption from the plausibility manipulation. This can help us to better understand whether ambiguous sentences encouraged longer rereading times overall – which would lead to longer rereading times on the downstream noun region following homographs than unambiguous words. Alternatively, readers may have targeted their rereading time specifically to re-examine just the interest areas that bore a direct relationship with the implausibility – in which case we would predict longer downstream noun rereading times for unambiguous sentences.

Rereading times for this region were calculated in the same way as those for the target word region, including a "0" value if the region was not returned to after the first pass. Rereading times on the downstream noun region were longer in *unambiguous* relative to ambiguous sentences, in addition to being longer for the implausible condition across both sentence types (Unambiguous Implausible: M=229 ms, SD=160; Unambiguous Plausible: M=128 ms, SD=95; Ambiguous Implausible: M=164 ms, SD=128; Ambiguous Plausible: M=119 ms; SD=102). Two 2×2 repeated-measures ANOVAs were constructed, again with plausibility (plausible vs. implausible) as a within-participant and within-items factor, and ambiguity of the preceding target (ambiguous vs. unambiguous) as a within-participants but between-items factor. Results confirmed a main effect of ambiguity, $F_1(1,23)=50.31$, $p<.05$, $F_2(1, 94)=10.46$, $p<.05$, plausibility, $F_1(1,23)=37.10$, $p<.05$, $F_2(1, 94)=82.76$, $p<.05$, as well as a significant interaction, $F_1(1,23)=10.80$, $p<.05$, $F_2(1, 94)=12.78$, $p<.05$. Follow-up t-tests found that the plausibility effect was significant within both the ambiguous sentences, $t_1(23)=5.75$, $p<.05$, $t_2(47)= 4.10$, $p<.05$), as well as the unambiguous sentences, $t_1(23)=5.26$, $p<.05$, $t_2(47)= 8.56$, $p<.05$), although numerically the magnitude of the plausibility effect in the unambiguous sentences (~100 ms) was more than double that in the ambiguous sentences (~46 ms).

Finally, we examined rereading times on the downstream noun region as a function of dominance of the preceding homograph (as can be seen in Figure 6). If the increased target region rereading times in the subordinate plausible condition were indicative of a generalized rereading strategy in those sentences, then we would expect that condition to again show the longest rereading times on the downstream noun region. Although implausible reading times were overall longer than plausible rereading times, they were actually the *shortest* in the subordinate-plausible cases (see Table 4 for rereading values). Analyses were conducted using 4 (dominance: dominant, subordinate, balanced, unambiguous) × 2 (plausibility: plausible, implausible) repeated measures ANOVAs, for which plausibility was a within-participant and within-item factor, and dominance was a within-participant but between-item factor. Results showed a main effect of dominance,

$F_1(3,69)=7.06$, $p<.05$, $F_2(3, 140)=5.01$, $p<.05$, a main effect of plausibility, $F_1(1,23)=42.89$, $p<.05$, $F_2(1, 140)=82.76$, $p<.05$, and a significant interaction between the two, $F_1(3,69)=6.40$, $p<.05$, $F_2(3, 140)=4.93$, $p<.05$. Follow-up t-tests examining the plausibility effect within each dominance condition found that the effect was significant following dominant instantiations, $t_1(23)=5.65$, $p<.05$, $t_2(32)= 4.28$, $p<.05$, reliable by participants only following balanced homographs, $t_1(23)=3.61$, $p<.05$, $t_2(29)= 1.82$, $p=.08$, and not significant following subordinate instantiations, $t_1(23)=1.12$, $p=.27$, $t_2(32)= 0.53$, $p=.61$. Comparing just the dominant and subordinate cases, we found that their plausible rereading times did not differ, $t_1(23)=0.88$, $p=.39$, $t_2(32)= 0.48$, $p=.63$, but that implausible rereading times were significantly *shorter* following the subordinate instantiations than the dominant, $t_1(23)=3.47$, $p<.05$, $t_2(32)= 2.78$, $p<.05$. Thus, rereading times on the downstream noun region disconfirm the possibility that readers generally reread more in the subordinate plausible condition. They instead suggest a more targeted strategy wherein readers devote the most time to rereading the area where the implausibility was first cued: the downstream noun region in the unambiguous sentences, and the target homograph region in the ambiguous sentences.

## Discussion

### First-Pass Ambiguity Effects

The current study investigated online lexical ambiguity resolution and its consequences for downstream processing. We expected to replicate our finding that ambiguous words show longer first fixation durations than unambiguous words when only syntactic cues are available for disambiguation (Stites et al., 2013), and indeed we did. First fixation durations to NV homographs were longer than those to unambiguous words when the preceding contexts contained syntactic cues to disambiguate the homograph and were semantically coherent, but not semantically constraining with respect to the intended meaning. These same contexts have previously elicited sustained frontal negativity effects in young adult readers (Lee & Federmeier, 2009, 2012). We furthermore predicted that, if the first fixation and frontal negativity effects share the same neurocognitive source, then they should be elicited in a similar timeframe and in the same regions. This prediction was also supported by the data: both effects are observed around 200 ms after the NV homographs were first apprehended, and first fixation durations are increased on both sentence regions in which the frontal negativity effect has been shown to be sustained. This study is now the second to show that first fixation durations to NV homographs are susceptible to influence from the same source that generates the frontal negativity effect, providing evidence that both effects likely reflect recruitment of the same frontally-mediated selection processes that are necessary in the face of difficult meaning selection.

Additionally, the dominance-based analyses found that the first fixation effect was numerically larger when the context instantiated the subordinate meaning of the NV homograph, and, indeed, only first fixations to the subordinate instantiations were significantly longer than those to unambiguous words (the comparison between the subordinate and dominant condition was also significant in the by-participants analysis). These results provide evidence that some meaning selection mechanisms likely involve the

*active inhibition* of context-inappropriate meanings, and that this inhibition signal slows the eyes on first fixation durations. The fact that first fixations on balanced words and the dominant meaning of biased words did not differ from those in the unambiguous condition suggests that other types of meaning selection can be done without the recruitment of these frontally-mediated active inhibition mechanisms. Similarly, readers can select even the subordinate meaning of NV homographs using semantic constraints without eliciting the frontal negativity effect (Lee & Federmeier, 2009). Instead, inhibition seems to be necessary to allow activation of a weak meaning in the face of a competitor, when cues to meaning selection come from outside the semantic system -- i.e., the circumstances in which we observe the first fixation and frontal negativity effects.

The lack of an ambiguity effect on gaze durations on the target word and the preposition region replicate our previous findings (Stites, et al., 2013) that first fixation durations are the most susceptible to influence from the mechanisms generating the frontal negativity ambiguity effect. Although first fixations and gaze durations can be closely related (indeed, are equivalent when targets are only fixated once), we interpret this dissociation in effects as suggesting that these two reading time measures are driven by at least partially non-overlapping processes. Because the frontal negativity effect likely involves inhibitory processing, it could reflect a “sweep” of inhibition through the language processing system. Based on the early onset of the frontal negativity effect (~200 ms post-stimulus onset), this general inhibitory processing could also briefly inhibit eye movements in the same time window (i.e., during first fixation durations). However, after the initial sweep through the system, other higher-level language comprehension processes will then continue to exert their influence over eye movement control, which could explain why the refixations to targets that constitute gaze durations (as well as those “first” fixations to the ambiguous words that were preceded by fixations to the immediately preceding function word and thus may actually be more like refixations to the target) do not show the same ambiguity effects.

### First-Pass Plausibility Effects

The core question of the current study involved characterizing plausibility effects at the downstream noun of the prepositional phrase, as a means to test the functional significance of the first fixation effect. The downstream noun served as a “probe word,” and was plausible given only one sense of the target homograph. Reading times to these nouns in the ambiguous (compared to unambiguous) condition thus served as a test of the extent to which the context-inappropriate meaning of the NV homograph was suppressed. If the first fixation effect reflects meaning *selection*, then its elicitation at the NV homograph should produce longer downstream reading times for implausible compared to plausible nouns. This would indicate that readers selected the context-appropriate meaning, thus making the context-inappropriate meaning less available and rendering the downstream noun associated with it implausible. If, on the other hand, the first fixation effect reflects meaning *maintenance*, then its elicitation at the NV homograph should produce equally facilitated reading times on the implausible and plausible downstream nouns, because both meanings would be actively held in memory.

The downstream noun of the prepositional phrase was not directly semantically associated with either meaning of the NV homograph. This feature of the current stimulus set helped avoid reactivation of the homograph's alternate meaning, which could cloud interpretation of facilitation effects observed on the probe (Van Petten and Kutas, 1987). Secondly, the matched levels of sentential constraint and cloze probability for the plausible and implausible words allowed us to improve upon the design of previous eye-tracking studies of plausibility effects, which appear to confound target plausibility with its predictability from the preceding context (Warren et al., 2008) because the plausible words are also more predictable than the implausible words. Others studies do not report the overall constraint of the sentences used or the cloze probability of their particular target words (e.g., Filik, 2008; Rayner et al., 2004; Warren & McConnell, 2007), making it unclear whether readers may have been actively predicting a different word when they received an implausible/impossible noun. ERP research has shown that readers engage in different types of processing when they receive an equally unexpected word depending on whether or not they were predicting a more likely word (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Wlotko & Federmeier, 2007), which could potentially translate into different patterns of eye movement effects. The current design allows us to test whether first-pass measures are indeed subject to plausibility effects in the absence of this potential confound. We also asked readers to engage in an overt task that was quite different from that used in other eye-tracking studies of plausibility: readers responded to a probe word at the end of every sentence, whereas other studies tend to ask fewer, more comprehension-based questions. The probe word task could affect how closely readers attend to the materials, which could in turn impact the timing and magnitude of their plausibility effects.

Although Lee and Federmeier (2012) found evidence linking the frontal negativity with meaning selection in their ERP study, it is possible that the effects could play out differently in an eye-tracking paradigm in which readers have more control over their intake of information from the sentence. Specifically, this control may encourage readers to maintain both meanings of the homograph until more information becomes available, because they know that they can return to earlier parts of the text if they are still confused later. In contrast, it is not possible to move back in the text in the RSVP presentation style necessitated by ERP research. This could instead encourage readers to adopt an early selection strategy, wherein they select a meaning immediately upon encountering the ambiguous word because they know they cannot return to it later. The present study thus allows an important test of the generalizability of the ERP findings to natural reading.

Our results showed that ambiguity effects observed in eye-tracking (first fixation effects) as in ERPs (frontal negativity effects) seem to arise from meaning selection processes: first-pass reading times were longer for implausible than plausible downstream noun regions following both NV homographs and unambiguous words. These plausibility effects were just as large following ambiguous as unambiguous words, suggesting that the processes generating the first fixation effect enabled readers to suppress the context-inappropriate meaning of the homograph well enough to render it implausible later in the sentence. These findings replicate Lee and Federmeier (2012) in showing that young adults elicit plausibility effects of roughly equal magnitude following ambiguous and unambiguous words. The use of the downstream plausibility manipulation thus allows us to draw an explicit connection

between the first fixation ambiguity effect and meaning selection mechanisms, which was not possible in our previous study (Stites et al., 2013). It is encouraging that we observed these plausibility effects despite the different task conditions in our study (probe word as opposed to comprehension questions), suggesting that these effects transcend the particulars of the experimental context. Furthermore, as will be discussed in more detail below, our results show that *implausibility*, not just *impossibility*, can produce effects on first-pass reading measures.

### Relationship Between Ambiguity and Plausibility Effects

If the first fixation effect reflects the recruitment of meaning selection mechanisms, as we have argued it does, then it should share the same relationship with the downstream plausibility effects as was observed in Lee and Federmeier (2012). Lee and Federmeier (2012) found a positive correlation between the size of a reader's frontal negativity effect and the size of his/her N400 plausibility effect following ambiguous words, suggesting that the context-inappropriate meaning of the NV homograph was less available downstream as a reader's frontal negativity effect at the ambiguous word increased in size. We found the same positive correlation between the size of a reader's first fixation effect on the NV homographs and the size of his/her downstream plausibility effect. In other words, the bigger one's first fixation effect, the less available the word's context-inappropriate meaning was at the downstream noun region, which produced larger plausibility effects. This finding strongly suggests that the first fixation effect is susceptible to influence from the underlying processes that afford meaning selection. The correlation was largest with the plausibility effect for go-past time on the noun region, which captures all reading times from a reader's first fixation on the noun region until they move past it to the right. As such, it includes initial reading times on both words of the region, as well as any time they spent rereading earlier parts of the sentence (including the earlier ambiguous word) before moving forward. That this measure, rather than first fixation or gaze duration, showed the strongest correlation suggests that readers actively allocated their attention to reading earlier parts of the sentence to resolve implausibility, rather than simply slowing down upon encountering the violating word. Importantly, we demonstrated that the correlation between these effects was specific to cases following ambiguous words but not following unambiguous words, and as such, was not due to a reader's general tendency to slow down on difficult sections of a text (i.e., all ambiguities *and* all implausibilities). The selective relationship between the first fixation ambiguity effect and plausibility effects following ambiguous words provides strong evidence for the involvement of meaning selection processes in the generation of this effect.

The strikingly similar findings between our current eye-tracking results and the ERP findings using the same stimuli (Lee and Federmeier, 2012) have established a compelling relationship between the first fixation and frontal negativity effects. Our previous work (Stites et al., 2013) showed that both effects are elicited by NV homographs in the presence of syntactic cues without biasing semantics, are reduced or absent in older adults as a group, and are positively correlated with a reader's verbal fluency. The current study adds to our understanding of their relationship by showing again that both are elicited by NV homographs in (different kinds of) semantically neutral contexts, both are then observed in

the following prepositional phrase, and both are correlated with the size of an individual's downstream plausibility effect. Together, the findings from this and our previous study point to the likelihood that these two effects share an underlying function and possibly even the same neural generators. Although the source of the frontal negativity effect cannot be directly assessed using ERPs alone, other imaging work points to the role of the left inferior frontal gyrus (LIFG) in ambiguity resolution (Bilenko et al., 2009; Gennari, MacDonald, Postle, & Sidenberg, 2007; Mason & Just, 2007; Zempleni, Renken, Hoeks, Hoogduin, & Stowe, 2007) and semantic selection more broadly (for reviews, see Novick, Trueswell, & Thompson-Schill, 2010, and Thompson-Schill, Bedny, & Goldberg, 2005).

### Plausibility Effects in Gaze Measures

The current study also gives us the opportunity to consider the relationship between plausibility effects in eye-tracking and ERPs more generally. Eye-tracking researchers typically examine plausibility effects in terms of which *word* of the sentence first shows increased reading times. This has resulted in classifying some effects as “immediate”, referring to increased first-pass measures on the violating word itself, and others as “delayed” effects, referring to increased first-pass measures on subsequent words *or* later reading time measures on the violating word. It has been suggested that the degree of the violation can affect the timing with which it affects the eye movement record, with *impossible* violations producing immediate eye-tracking effects and *implausible* violations largely producing delayed effects (Warren, 2011). Interpreting plausibility effects in eye-tracking studies becomes especially tricky when we consider that readers have a range of options for how to continue when they reach an implausible word. For example, they could slow down on the violating word, move forward in the text (and possibly slow down there), or immediately regress back to an earlier point of the sentence. The choice of how to proceed may vary from reader to reader, or trial to trial, and, as a result, it is unclear which word in the sentence, or even which reading time measure, provides the “best” representation of plausibility effects. This may contribute to the inconsistent characterizations of plausibility effects across the eye-tracking literature.

Work using ERPs suggests that it might be more fruitful to focus on the *timepoint* at which plausibility effects are observed rather than looking at reading times on any particular word. ERP work has consistently found that N400 plausibility effects are elicited by the violating word itself (Ferguson, Sanford, & Leuthold, 2008; Filik & Leuthold, 2008; Lee & Federmeier, 2012; Nieuwland, 2013; Nieuwland & Kuperberg, 2008; Nieuwland & Martin, 2012; Nieuwland & Van Berkum, 2006; Paczynski & Kuperberg, 2012), and such N400 effects are always maximal approximately 300-500 ms after readers apprehend the implausible word. We would thus predict that the effects of plausibility would be observed along a similar timecourse in natural reading as well, regardless of where readers’ eyes are fixated at that point in time. Our approach in the current study was to apply this prediction from ERP work to our analysis strategy by combining the downstream noun with its subsequent spillover word to make a region. In this way, we can capture readers’ processing of the implausible target words as it unfolds over time, even when they have moved their eyes forward in the text. Although this approach will allow us to better align the timelines between eye-tracking and ERP effects, one caveat that remains is that with eye-tracking it is

not possible to pinpoint the exact moment that readers first began processing the noun, due to the (potential) availability of preview information (which is not available in typical ERP studies). As such, any reading time measure on the critical word will still likely underestimate the full amount of time readers spent processing that word, because it will be unable to account for any time that readers might have spent gathering information about the word during their previous fixation. Regardless of this shortcoming, when we defined the time window following the apprehension of the downstream noun in this way, we saw effects of implausibility on first pass reading times to the region (which have been rare in eye-tracking studies with *plausibility* violations). Interestingly, average gaze durations on the noun region fell between 385-429 ms (see Table 3), which lines up nicely with the time frame in which N400 plausibility effects are observed, for these materials as in other studies. We believe that our findings represent a more general relationship between the timecourse of ERP and eye-tracking effects -- particularly that effects expressed as modulations of the N400 will often be “smeared” across the reading of several words, due to the extensive amount of saccadic pre-programming hypothesized to underlie skilled reading in multiple models of eye movement control (Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Pollatsek, & Rayner, 2006). Because readers cannot often recognize that a word is implausible fast enough to affect reading times on the word itself, they will likely continue forward in the text before the semantics of that word can be appreciated in light of the ongoing message-level context. As such, the increased reading times associated with plausibility violations may not show up until the following word (or even two regions later in the sentence, as has been observed in a self-paced reading study [Matsuki et al, 2011])[2]. On the other hand, if a lot of readers choose to look back in the text, either due to individual variation across samples or characteristics of the experimental stimuli, then the plausibility effects might be captured instead in regression out probabilities rather than spillover reading times (as observed by Warren & McConnell, 2007). In the current study, the percentage of first pass regressions out from the downstream noun was low (12-14%, with no effects of plausibility or ambiguity). The vast majority of the time, then, readers moved directly from the downstream noun to the spillover word immediately following it, where their effects were captured in first-pass noun region reading times.

We are not the first to suggest that ERP effects, particularly those on the N400, may be spread across multiple fixations in natural reading. Dambacher and Kliegl (2007) suggested a similar relationship between the timing of eye movement and ERP effects. They compared reading times and ERPs elicited by the same group of sentences (recorded at different times) containing words of varying frequencies. Their results showed that low-frequency words elicited larger N400 amplitudes than high frequency words (a typical N400 frequency effect), as well as longer reading times on the *next* word in the sentence. Both sets of findings strongly suggest that the processing of a single word likely continues even after readers are no longer fixating it. This processing can influence subsequent fixations at approximately the same timepoint as the effects would be observed on the N400, especially

[2]Studies employing possibility violations (Warren & McConnell, 2007; Warren, McConnell, & Rayner, 2008), which often report earlier reading time effects, used words that differed greatly from the most expected word. Thus, parafoveal preview information may have allowed readers to begin to recognize its incongruence with the context on the fixation prior to landing on it. This would allow processing of the impossible word to begin while readers were fixated on the previous word and, in line with our proposed timeline of plausibility effects, affect reading times on the target word itself.

when the word is difficult due to its low frequency or implausibility with the preceding context.

### Rereading Time Effects

In order to assess whether readers experienced continued difficulty with either the homographs or the implausible downstream nouns after the first pass, we examined rereading times on each of these regions. We combined the target homograph with its preceding function word into a region, because either word (or both) would make a reasonable target to return to in order to resolve residual uncertainty as to the intended meaning of the target word. Multiple large-scale patterns emerge when we consider rereading times on the target word region in tandem with those to the downstream noun region. First, in the *unambiguous* sentences, readers showed a large plausibility effect on the downstream noun region (i.e. spent more time rereading the implausible than plausible nouns), but did not show a similar effect on the target homograph region. For example, in the implausible sentence "Karen / expected / the member / of the / **mother** to / bring his own instrument," readers spent more time rereading the implausible downstream noun region "***mother*** *to"* than they did rereading the unambiguous target region "*the member*." This is likely because the downstream noun region was the first point at which the sentence was implausible, and there was no systematic relationship between the target "*member*" and the implausible continuation "***mother***".

Secondly, rereading times on the target homograph region were longer for ambiguous than unambiguous targets, especially so in the implausible sentences. Why might readers return to the *homograph* when the *downstream noun* is the implausible word? We propose that the elicitation of the first fixation effect at the NV homograph generated an error signal that "marked" that word as a place of uncertainty. Readers may have then chosen to return to this region as part of a strategy in which they revisited the site of a previous uncertainty, even if the ambiguity itself was resolved in the first pass. This falls in line with a similar proposal by Levy and colleagues (Levy, Bicknell, Slattery, & Rayner, 2010) that readers maintain uncertainty about word identities, especially of short function words, after moving past them, and they can use this uncertainty to selectively return to that word when they later encounter an ambiguity whose resolution depends on its interpretation. With respect to the current study, the implausibility of our ambiguous sentences hinges on the interpretation of the homograph, because the downstream noun was plausible with the context-*inappropriate* instantiation. Because the correct meaning of the homograph is instantiated through a single, short function word (*to* or *the*), it seems reasonable that readers would choose to return to it after encountering a later implausibility as a way to "double check" that they correctly interpreted the original homograph.

This pattern is exactly what we found in the implausible sentences containing homographs: readers spent much more time rereading the target homograph region when the downstream noun was implausible. For example, in the implausible sentence "Danny / preferred / the scale / with the / **knife** that / also provided a body mass index," rereading times were much longer on the target region "*the scale*" in the implausible than plausible condition (in which the downstream word was ***display***). The fact that readers elicit equally long implausible

rereading times for the subordinate instantiations indicates that readers were indeed able to inhibit the dominant meaning – otherwise the downstream implausible word would have been judged as plausible with the still-active dominant meaning, thus reducing the need to return to the target region for re-analysis.

Interestingly, though, rereading times in the subordinate *plausible* condition were just as long as those in the implausible condition, a pattern that was unique to the subordinate condition. In other words, in the plausible sentence "Danny / preferred / to scale / with the / **knife** that / his grandfather had used for fishing," readers spent just as much time rereading "*to scale*" when it was followed by the plausible continuation "*knife*" as when it was followed by the implausible continuation "*display*." This pattern of rereading times could have two related interpretations. First, as previously discussed, it could indicate uncertainty by the readers about the correct instantiation of the target word, and/or of the identity of its preceding function word. This effect could be magnified when readers must select the subordinate meaning of the homograph, causing them to return to the subordinate target irrespective of the plausibility of the context that followed it. A second interpretation is that this effect reflects incomplete meaning selection of the subordinate sense of the homograph. Residual activation of the dominant meaning could have persisted even though the selection processes indexed by the first fixation effect were at least effective enough to generate the downstream plausibility effects. This lingering activation could have caused readers to incorrectly judge the plausible subordinate-related downstream noun as being implausible (at least some of the time), thus spurring more rereading of the target word region in these cases.

## Conclusions

In summary, this study found the predicted increased first fixation durations on NV homographs when only syntactic cues were available, an effect that was also present on first fixations to the preposition region. This effect is elicited by the same conditions and in the same sentence regions as the frontal negativity effect in Lee and Federmeier (2012), and replicates our previous findings of increased first fixation durations to NV homographs in semantically incoherent but syntactically constraining contexts (Stites et al., 2013). Paralleling the N400 plausibility effects found by Lee and Federmeier (2012) on the downstream noun, we also found downstream plausibility effects in eyegaze measures, of equal size following ambiguous and unambiguous words. Moreover, like Lee and Federmeier (2012), we found a positive correlation between the ambiguity effect and the downstream plausibility effect. Together, these findings constitute strong evidence that readers bring selection mechanisms online to effect difficult ambiguity resolution, resulting in the suppression of context-irrelevant meanings of ambiguous words. More generally, the results underscore the benefits of seeking converging evidence from multiple methods in the study of reading processes.

## Acknowledgements

This research was supported by National Institutes of Health (NIH) Predoctoral Training Grant T32-HD055272 and National Science Foundation Graduate Research Fellowship DGE 11-44245 FLLW to Mallory C. Stites, and a

James S. McDonnell Foundation Scholar Award and NIH Grant AG026308 to Kara D. Federmeier. The authors would like to acknowledge Chia-lin Lee for her contributions to stimulus development.

# Appendix

## Appendix

Additional example stimuli used in the experiment are listed below. For the ambiguous sentences containing Noun/Verb Homographs (Word Type *NVH*), target homographs are italicized, the dominant sense is underlined (where applicable), and the downstream noun is bolded. For the unambiguous items (Word Type *UW*), the column labeled *Dominant Sense* lists the control word's syntactic class.

| Word Type | Target | Dominant Sense | Plausibility | Sentence |
|---|---|---|---|---|
| NVH | Bat | Verb | Plausible | Josh liked to *bat* in the **game** when the bases were loaded. |
| | | | | Kevin liked the *bat* in the **cave** but knew that it would upset his wife. |
| | | | Implausible | Josh liked the *bat* in the **game** when the bases were loaded. |
| | | | | Kevin liked to *bat* in the **cave** but knew that it would upset his wife. |
| NVH | Date | Noun | Plausible | Janice needed to *date* in the **evening** because she got off work really late. |
| | | | | Helen needed the *date* in the **calendar** to be marked for her appointment. |
| | | | Implausible | Janice needed the *date* in the **evening** because she got off work really late. |
| | | | | Helen needed to *date* in the **calendar** to be marked for her appointment. |
| NVH | Drill | Verb | Plausible | Jorge learned to *drill* through the **wood** using slow, even pressure. |
| | | | | Greg learned the *drill* through the **bootcamp** when he first got into the military school. |
| | | | Implausible | Jorge learned the *drill* through the **wood** using slow, even pressure. |
| | | | | Greg learned to *drill* through the **bootcamp** when he first got into the military school. |
| NVH | Fly | Verb | Plausible | Katherine wanted to *fly* on the **plane** rather than drive the long distance. |
| | | | | Jamie wanted the *fly* on the **wall** to go away. |
| | | | Implausible | Katherine wanted the *fly* on the **plane** rather than drive the long distance. |
| | | | | Jamie wanted to *fly* on the **wall** to go away. |
| NVH | Litter | Noun | Plausible | Robert prepared to *litter* in the **street** until he saw the police officer on the corner. |
| | | | | Tom prepared the *litter* in the **box** by raking it out the way the cat liked best. |

| Word Type | Target | Dominant Sense | Plausibility | Sentence |
|---|---|---|---|---|
| | | | Implausible | Robert prepared the *litter* in the **street** until he saw the police officer on the corner.<br>Tom prepared to *litter* in the **box** by raking it out the way the cat liked best. |
| NVH | Notice | Balanced | Plausible | Tina forgot to *notice* on her **way** whether the new store was open yet.<br>Kerry forgot the *notice* on her **door** and wasn't there for the package redelivery. |
| | | | Implausible | Tina forgot the *notice* on her **way** whether the new store was open yet.<br>Kerry forgot to *notice* on her **door** and wasn't there for the package redelivery. |
| NVH | Rule | Balanced | Plausible | Sharon wanted to *rule* in the **kingdom** as the supreme and benevolent monarch.<br>Kelly wanted the *rule* in the **handbook** to be more specific. |
| | | | Implausible | Sharon wanted the *rule* in the **kingdom** as the supreme and benevolent monarch.<br>Kelly wanted to *rule* in the **handbook** to be more specific. |
| NVH | Sink | Balanced | Plausible | Roland wanted to *sink* in his bed to take a long nap.<br>Ian wanted the *sink* in his **bathroom** to be repaired. |
| | | | Implausible | Roland wanted the *sink* in his **bed** to take a long nap.<br>Ian wanted to *sink* in his **bathroom** to be repaired. |
| NVH | Worm | Noun | Plausible | Her brother wanted to *worm* out of the **chore** but did the dishes anyway.<br>His nephew wanted the *worm* out of the **apple** because he thought it unsanitary. |
| | | | Implausible | Her brother wanted the *worm* out of the **chore** but did the dishes anyway.<br>His nephew wanted to *worm* out of the **apple** because he thought it unsanitary. |
| UW | Beauty | Noun | Plausible | Joan was captivated by the *beauty* of the **gorge** and decided to come back again. |
| | | | Implausible | Joan was captivated by the *beauty* of the **skills** and decided to come back again. |
| UW | Behave | Verb | Plausible | Eva needed to *behave* during the **mass** so her parents wouldn't be mad. |
| | | | Implausible | Eva needed to *behave* during the **road** so her parents wouldn't be mad. |
| UW | Come | Verb | Plausible | Eric wanted to *come* from the **airport** directly on the day he flies in. |

| Word Type | Target | Dominant Sense | Plausibility | Sentence |
|---|---|---|---|---|
| | | | Implausible | Eric wanted to *come* from the **subjects** directly on the day he flies in. |
| UW | Income | Noun | Plausible | Daniel needed the *income* from this **contract** to pay for the mortgage. |
| | | | Implausible | Daniel needed the *income* from this **dark** to pay for the mortgage |
| UW | Lend | Verb | Plausible | Karen refused to *lend* to this **person** because of his poor credit history. |
| | | | Implausible | Karen refused to *lend* to this **railing** because of his poor credit history. |
| UW | Mile | Noun | Plausible | Tory runs the *mile* around the **neighborhood** as part of her daily exercise. |
| | | | Implausible | Tory runs the *mile* around the **deadline** as part of her daily exercise. |
| UW | Ocean | Noun | Plausible | Anita likes the *ocean* under the **moonlight** during summertime. |
| | | | Implausible | Anita likes the *ocean* under the **computer** during summertime. |
| UW | Remove | Verb | Plausible | He had to *remove* from his **collection** the paintings that were fakes. |
| | | | Implausible | He had to *remove* from his **industry** the paintings that were fakes. |

## References


Bilenko NY, Grindrod CM, Myers EB, Blumstein SE. Neural correlates of semantic competition during processing of ambiguous words. Journal of Cognitive Neuroscience. 2009; 21(5):960–975. doi:10.1162/jocn.2009.21073.Neural. [PubMed: 18702579]

Coulson S, Kutas M. Getting it: human event-related brain response to jokes in good and poor comprehenders. Neuroscience Letters. 2001; 316:71–74. [PubMed: 11742718]

Dambacher M, Kliegl R. Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. Brain Research. 2007; 1155:147–62. doi:10.1016/j.brainres.2007.04.027. [PubMed: 17499223]

Drieghe D, Rayner K, Pollatsek A. Eye movements and work skipping during reading revisited. Journal of Experimental Psychology: Human Perception and Performanc. 2005; 31(5):954–969.

Engbert R, Nuthmann A, Richter EM, Kliegl R. SWIFT: A dynamical model of saccade generation during reading. Psychological Review. 2005; 112(4):777–813. doi:10.1037/0033-295X.112.4.777. [PubMed: 16262468]

Federmeier KD, Segal JB, Lombrozo T, Kutas M. Brain responses to nouns, verbs and class-ambiguous words in context. Brain : A Journal of Neurology. 2000; 123(Pt 12):2552–66. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11099456. [PubMed: 11099456]

Federmeier KD, Wlotko EW, De Ochoa-Dewald E, Kutas M. Multiple effects of sentential constraint on word processing. Brain Research. 2007; 1146:75–84. doi:10.1016/j.brainres.2006.06.101. [PubMed: 16901469]

Ferguson HJ, Sanford AJ, Leuthold H. Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. Brain Research. 2008; 1236:113–25. doi:10.1016/j.brainres.2008.07.099. [PubMed: 18722356]

Filik R. Contextual override of pragmatic anomalies: evidence from eye movements. Cognition. 2008; 106(2):1038–46. doi:10.1016/j.cognition.2007.04.006. [PubMed: 17524387]

Filik R, Leuthold H. Processing local pragmatic anomalies in fictional contexts: evidence from the N400. Psychophysiology. 2008; 45(4):554–8. doi:10.1111/j.1469-8986.2008.00656.x. [PubMed: 18282200]

Fitzsimmons G, Drieghe D. How fast can predictability influence word skipping during reading? Journal of Experimental Psychology: Learning, Memory, and Cognition. 2013; 39(4):1054–63. doi:10.1037/a0030909.

Folk JR, Morris RK. Effects of syntactic category assignment on lexical ambiguity resolution in reading: an eye movement analysis. Memory & Cognition. 2003; 31(1):87–99. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12699146. [PubMed: 12699146]

Frazier L, Rayner K. Resolution of syntactic category ambiguities : Eye movements parsing lexically ambiguous sentences. Journal of Memory and Language. 1987; 526:505–526.

Gordon PC, Plummer P, Choi W. See Before You Jump: Full Recognition of Parafoveal Words Precedes Skips During Reading. Journal of Experimental Psychology. Learning, Memory, and Cognition. 2012 doi:10.1037/a0028881.

Henry JD, Crawford JR. A meta-analytic review of verbal fluency performance in patients with traumatic brain injury. Neuropsychology. 2004; 18(4):621–8. doi:10.1037/0894-4105.18.4.621. [PubMed: 15506829]

Kutas, M.; King, JW. The potentials for basic sentence processing: Differentiating integrative processes.. In: Innue, T.; McClelland, J., editors. Attention and Performance XVI: Information Integration in Perception and Communication. MIT Press; Cambridge, MA: 1996. 1996

Lee C-L, Federmeier KD. To mind the mind: an event-related potential study of word class and semantic ambiguity. Brain Research. 2006; 1081(1):191–202. doi:10.1016/j.brainres.2006.01.058. [PubMed: 16516169]

Lee C-L, Federmeier KD. Wave-ering: An ERP study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. Journal of Memory and Language. 2009; 61(4):538–555. doi:10.1016/j.jml.2009.08.003. [PubMed: 20161361]

Lee C-L, Federmeier KD. Differential age effects on lexical ambiguity resolution mechanisms. Psychophysiology. 2011; 48(7):960–72. doi:10.1111/j.1469- 8986.2010.01158.x. [PubMed: 21175671]

Lee C-L, Federmeier KD. Ambiguity's Aftermath: How age differences in resolving lexical ambiguity affect subsequent comprehension. Neuropsychologia. 2012; 50(5):869–879. [PubMed: 22321956]

Matsuki K, Chow T, Hare M, Elman JL, Scheepers C, McRae K. Event-based plausibility immediately influences on-line language comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2011; 37(4):913–34. doi:10.1037/a0022964.

Nieuwland MS. “If a lion could speak …”: Online sensitivity to propositional truth-value of unrealistic counterfactual sentences. Journal of Memory and Language. 2013; 68(1):54–67. doi:10.1016/j.jml.2012.08.003.

Nieuwland MS, Kuperberg GR. When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. Psychological Science. 2008; 19(12):1213–8. doi:10.1111/j.1467-9280.2008.02226.x. [PubMed: 19121125]

Nieuwland MS, Martin AE. If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension. Cognition. 2012; 122(1):102–9. doi:10.1016/j.cognition.2011.09.001. [PubMed: 21962826]

Nieuwland MS, Van Berkum JJA. When peanuts fall in love: N400 evidence for the power of discourse. Journal of Cognitive Neuroscience. 2006; 18(7):1098–111. doi:10.1162/jocn.2006.18.7.1098. [PubMed: 16839284]

Novick JM, Trueswell JC, Thompson-Schill SL. Broca's area and language processing: Evidence for the Cognitive Control Connection. Language and Linguistics Compass. 2010; 4(10):906–924.

Paczynski M, Kuperberg GR. Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. Journal of Memory and Language. 2012; 67(4):426–448. doi: 10.1016/j.jml.2012.07.003. [PubMed: 23284226]

Patson ND, Warren T. Eye movements when reading implausible sentences: Investigating potential structural influences on semantic integration. Quarterly Journal of Experimental Psychology. 2011; 63(8):1516–1532. doi:10.1080/17470210903380822.Eye.

Princeton University "About WordNet.". WordNet. Princeton University; 2010. <http://wordnet.princeton.edu>

Rayner K, Frazier L. Selection mechanisms in reading lexically ambiguous words. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1989; 15(5):779–90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2528602.

Rayner K, Pacht JM, Duffy SA. Effects of prior encounter and global discourse on the processing of lexically ambiguous words: Evidence from eye fixations. Journal of Memory and Language. 1994; 33:527–544.

Rayner K, Slattery TJ, Drieghe D, Liversedge SP. Eye movements and word skipping during reading: effects of word length and predictability. Journal of Experimental Psychology. Human Perception and Performance. 2011; 37(2):514–28. doi:10.1037/a0020990. [PubMed: 21463086]

Rayner K, Warren T, Juhasz BJ, Liversedge SP. The effect of plausibility on eye movements in reading. Journal of Experimental Psychology. Learning, Memory, and Cognition. 2004; 30(6): 1290–301. doi:10.1037/0278-7393.30.6.1290.

Rayner K, Well AD. Effects of contextual constraint on eye movements in reading: A further examination. Psychonomic Bulletin & Review. 1996; 3(4):504–509. doi:10.3758/BF03214555. [PubMed: 24213985]

Reichle ED, Pollatsek A, Rayner K. E-Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. Cognitive Systems Research. 2006; 7:4–22.

Reichle ED, Warren T, McConnell K. Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. Psychonomic Bulletin & Review. 2009; 16(1):1–21. [PubMed: 19145006]

Rodd J, Gaskell G, Marslen-Wilson W. Making sense of semantic ambiguity: Semantic competition in lexical access. Journal of Memory and Language. 2002; 46(2):245–266. doi:10.1006/jmla.2001.2810.

Stuss DT, Levine B. Adult clinical neuropsychology: lessons from studies of the frontal lobes. Annual Review of Psychology. 2002; 53:401–33. doi:10.1146/annurev.psych.53.100901.135220.

Stites MC, Federmeier KD, Stine-Morrow EAL. May 20). Cross-age comparisons reveal multiple strategies for lexical ambiguity resolution during natural reading. Journal of Experimental Psychology: Learning, Memory, and Cognition. Advance online publication. 2013 doi: 10.1037/a0032860.

Thompson-Schill SL, Bedney M, Goldberg RF. The frontal lobes and the regulation of mental activity. Current Opinion in Neurobiology. 2005; 15:219–224. [PubMed: 15831406]

Warren, T. The influence of implausibility and anomaly on eye movements during reading.. In: Liversedge, S.; Gilchrist, I.; Everling, S., editors. The Oxford Handbook of Eye Movements. Oxford University Press; New York: 2011.

Warren T, McConnell K. Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. Psychonomic Bulletin & Review. 2007; 14(4): 770–5. [PubMed: 17972747]

Warren T, McConnell K, Rayner K. Effects of context on eye movements when reading about possible and impossible events. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2008; 34(4):1001–10. doi:10.1037/0278-7393.34.4.1001.

Wlotko EW, Federmeier KD. Finding the right word: Hemispheric asymmetries in the use of sentence context information. Neuropsychologia. 2007; 45(13):3001–14. doi:10.1016/j.neuropsychologia.2007.05.013. [PubMed: 17659309]

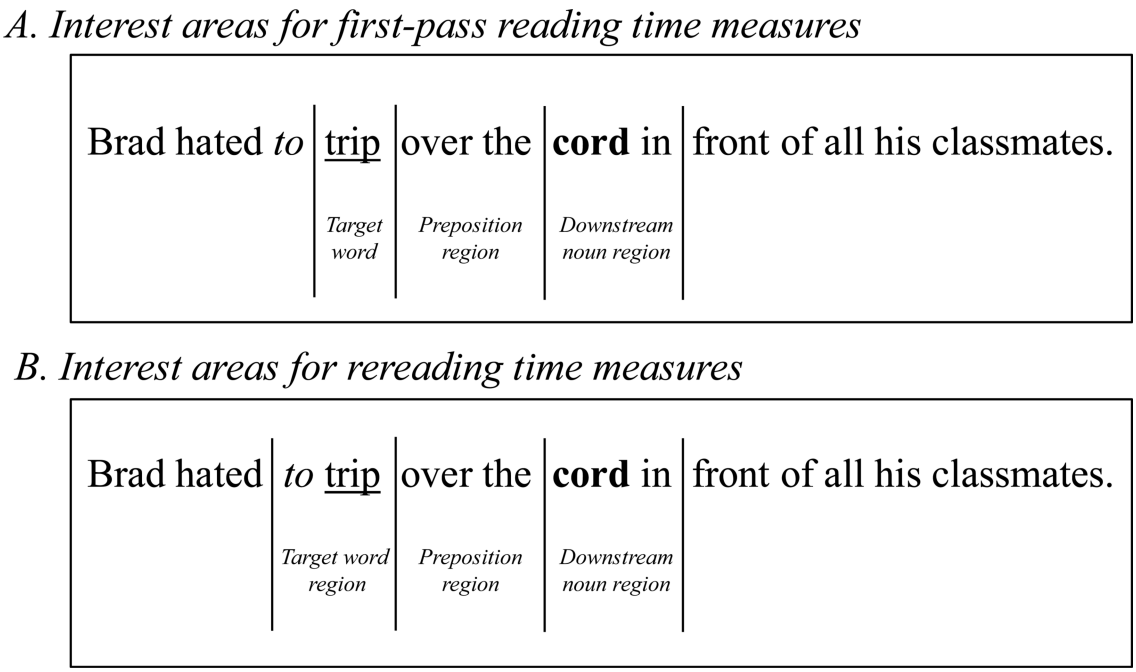


**Figure 1.**
Graphical representation of target sentence interest areas. A. Interest areas used for first-pass measures. B. Interest areas used for re-reading time measures (note that the target word is combined with the preceding determiner [*to* or *the*] to become the target word region).

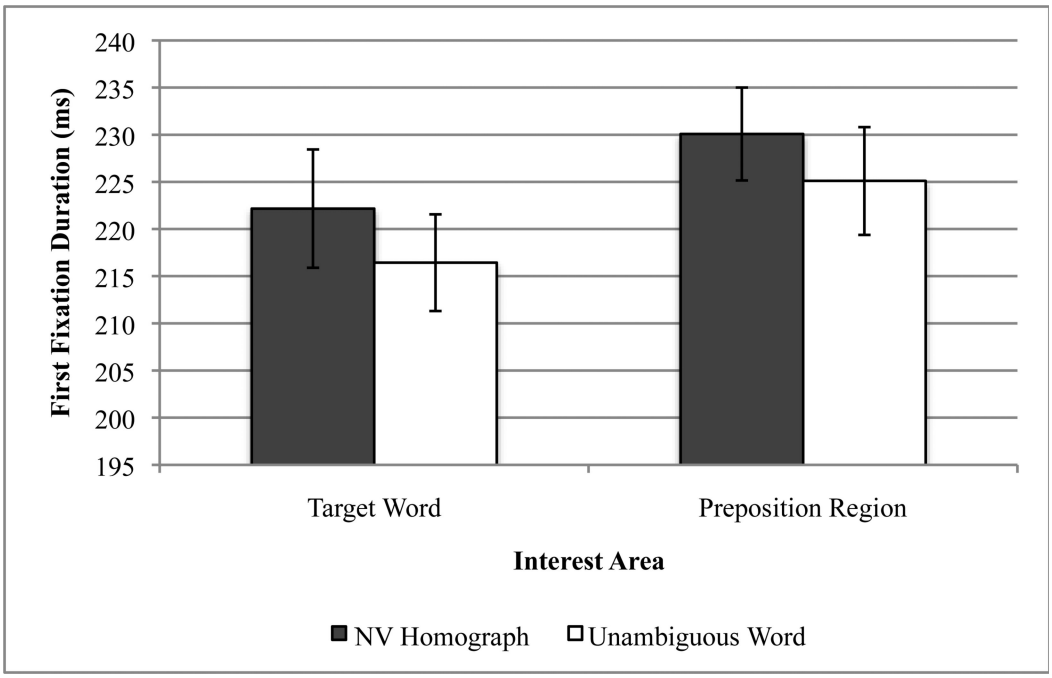


**Figure 2.**
First fixation durations on the target word and preposition region.

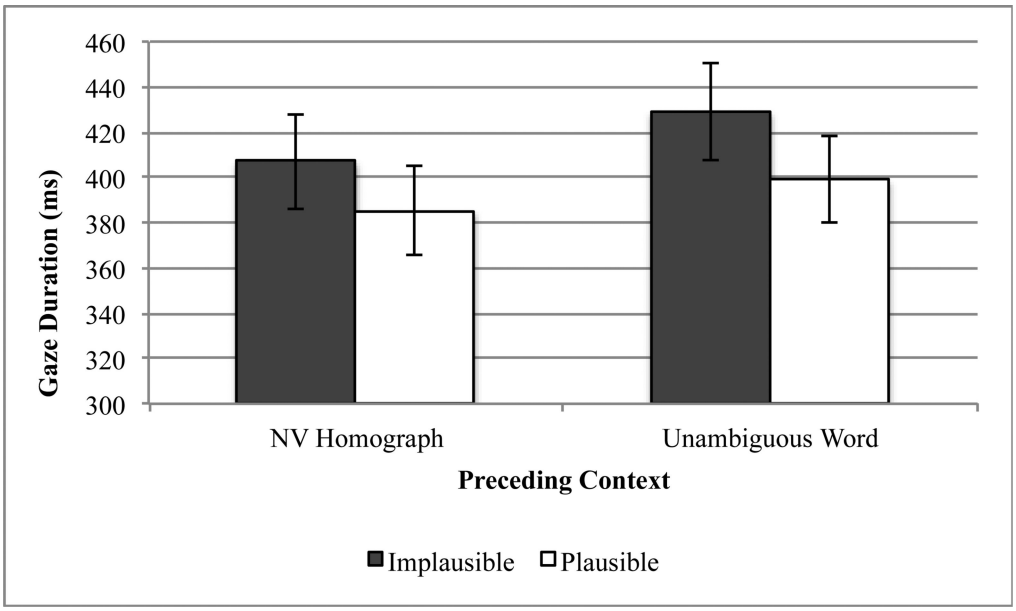


**Figure 3.**
Gaze duration on the noun region, split by the ambiguity of the preceding context.

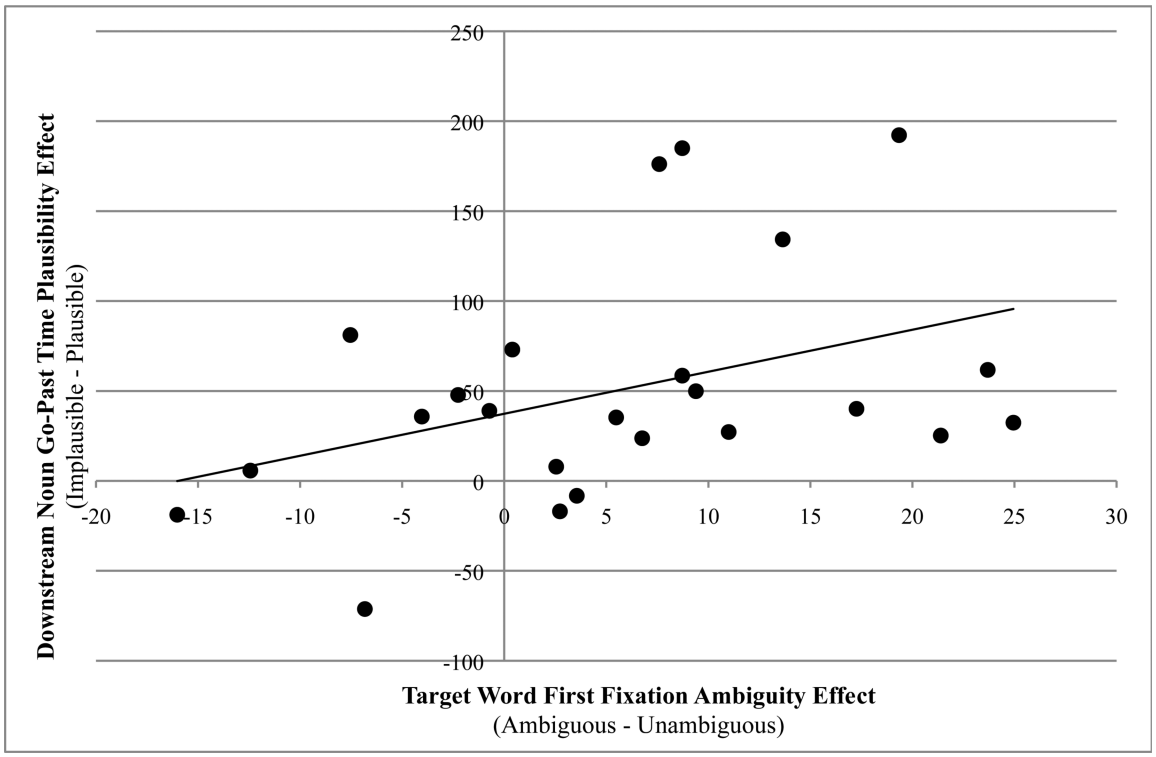


**Figure 4.**
Correlation between target word first fixation ambiguity effect and downstream noun region plausibility effect for go-past time.

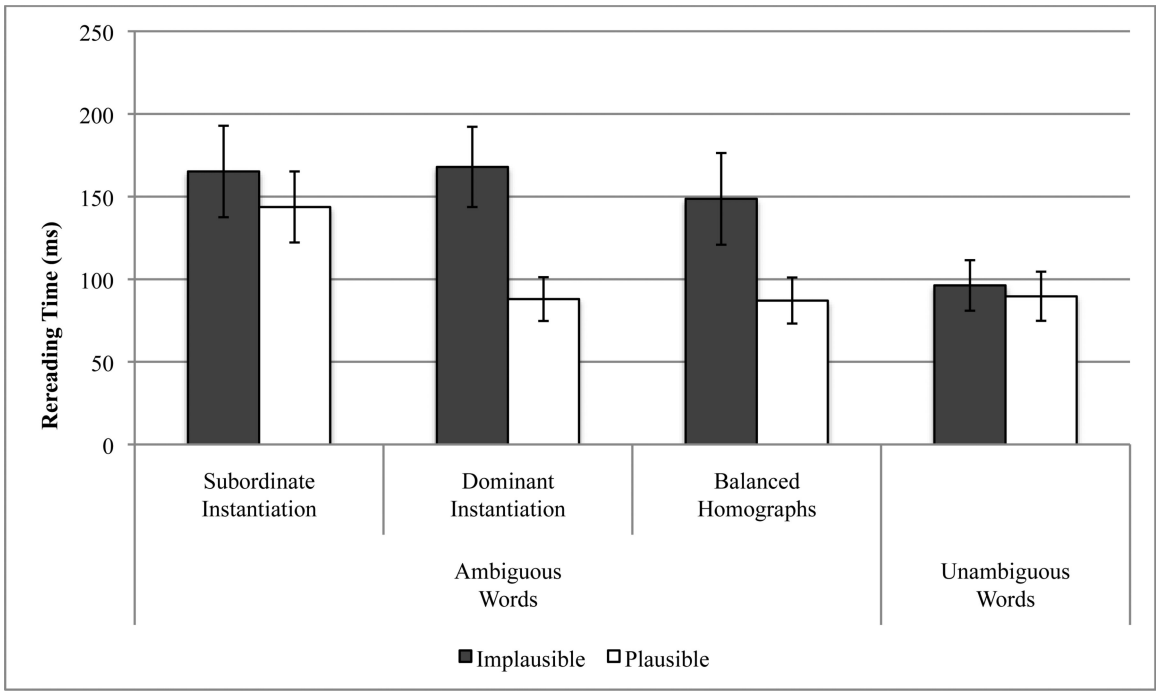


**Figure 5.**
Target word region rereading time, split by Target Word Dominance.

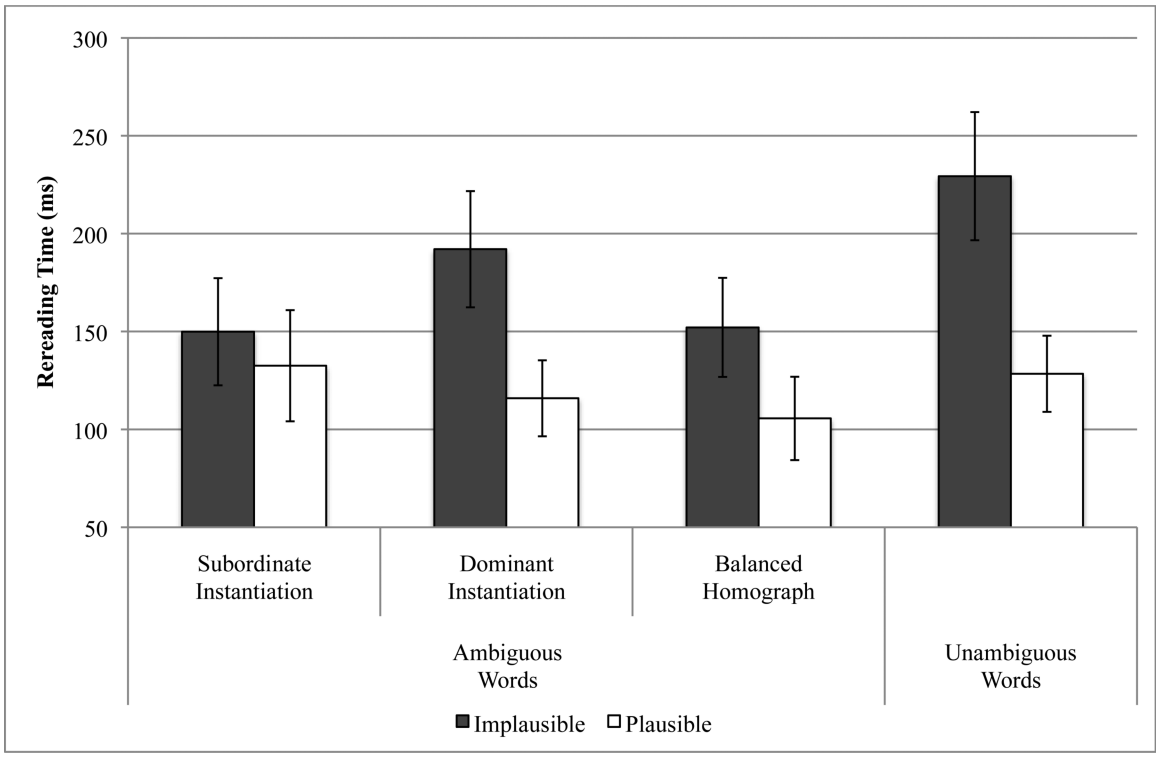


**Figure 6.**
Downstream noun region rereading time, split by Target Word Dominance.

**Table 1**

Mean values of lexical characteristics of target homographs and matched unambiguous words (standard deviations in parentheses).

| Measure | Unambiguous Words | Noun/Verb Homographs |
|---|---|---|
| Log frequency | 1.6 (0.5) | 1.6 (0.5) |
| Word length | 5.3 (1.0) | 4.5 (1.1) |
| Concreteness (1=very abstract; 7 = very concrete) | 4.8 (0.9) | 5.0 (1.0) |
| Semantic distinctiveness (1 = very different; 7 = very similar) | N/A | 2.5 (0.8) |

## Table 2

Examples of critical stimuli used in the experiment.

| Word Type | Plausibility | Example Sentence |
|---|---|---|
| **NV Homographs** | | *Noun-Dominant* |
| | Plausible | Brad hated to *trip* over the **cord** in front of all his classmates.<br>Brian hated the *trip* over the **river** because he hated getting wet. |
| | Implausible | Brad hated the *trip* over the **cord** in front of all his classmates.<br>Brian hated to *trip* over the **river** because he hated getting wet. |
| | | *Verb-Dominant* |
| | Plausible | Carrie hoped to *pass* on her **wisdom** to her grandchildren.<br>Laura hoped the *pass* on her **desk** would still be there when she returned home. |
| | Implausible | Carrie hoped the *pass* on her **wisdom** to her grandchildren.<br>Laura hoped to *pass* on her **desk** would still be there when she returned home. |
| | | *Balanced Homograph* |
| | Plausible | Ryan tried to *duck* in the **alley** to avoid the paparazzi.<br>Ben tried the *duck* in the **dish** prepared by a famous chef. |
| | Implausible | Ryan tried the *duck* in the **alley** to avoid the paparazzi.<br>Ben tried to *duck* in the **dish** prepared by a famous chef. |
| **Unambiguous Words** | Plausible | Joseph started to *appear* in the show since last season.<br>Karen expected the *member* of the band to bring his own instrument. |
| | Implausible | Joseph started to *appear* in the mistakes since last season.<br>Karen expected the *member* of the mother to bring his own instrument. |

## Table 3

First-pass reading measures (and standard deviations) for the three regions of interest.

| Region | Ambiguity | First Fixation | Gaze Duration | Go-Past Time |
|---|---|---|---|---|
| Target Word | A | 222 (31) | 252 (41) | 309 (56) |
| | U | 216 (25) | 256 (42) | 297 (53) |
| Preposition Region | A | 230 (24) | 323 (63) | 390 (93) |
| | U | 225 (28) | 322 (66) | 369 (84) |

| | | Implaus | Plaus | Implaus | Plaus | Implaus | Plaus |
|---|---|---|---|---|---|---|---|
| Downstream Noun Region | A | 226 (25) | 218 (18) | 407 (104) | 385 (97) | 527 (139) | 478 (124) |
| | U | 228 (28) | 221 (23) | 429 (104) | 399 (95) | 551 (155) | 473 (106) |

For the Ambiguity factor, A=Ambiguous Words and U=Unambiguous Words.

**Table 4**

Target Region and Downstream Noun Region Rereading times (and standard deviations) split by target word dominance.

| Region | Dominance | Rereading Time | |
|---|---|---|---|
| | | Implausible | Plausible |
| Target Region | Dominant | 168 (119) | 88 (65) |
| | Subordinate | 165 (136) | 144 (105) |
| | Balanced | 149 (87) | 87 (68) |
| | Unambiguous | 96 (75) | 90 (73) |
| Downstream Noun Region | Dominant | 192 (146) | 116 (95) |
| | Subordinate | 150 (134) | 133 (139) |
| | Balanced | 152 (124) | 106 (104) |
| | Unambiguous | 229 (160) | 128 (95) |