Published in final edited form as:
J Mem Lang. 2012 November 1; 67(4): 407–425. doi:10.1016/j.jml.2012.07.010.

Polysemy in Sentence Comprehension: Effects of Meaning Dominance

Stephani Foraker and
SUNY College at Buffalo

Gregory L. Murphy
New York University

Abstract

Words like *church* are polysemous, having two related senses (a building and an organization). Three experiments investigated how polysemous senses are represented and processed during sentence comprehension. On one view, readers retrieve an underspecified, core meaning, which is later specified more fully with contextual information. On another view, readers retrieve one or more specific senses. In a reading task, context that was neutral or biased towards a particular sense preceded a polysemous word. Disambiguating material consistent with only one sense followed, in a second sentence (Experiment 1) or the same sentence (Experiments 2 & 3). Reading the disambiguating material was faster when it was consistent with that context, and dominant senses were committed to more strongly than subordinate senses. Critically, following neutral context, the continuation was read more quickly when it selected the dominant sense, and the degree of sense dominance partially explained the reading time advantage. Similarity of the senses also affected reading times. Across experiments, we found that sense selection may not be completed immediately following a polysemous word but is completed at a sentence boundary. Overall, the results suggest that readers select an individual sense when reading a polysemous word, rather than a core meaning.

Keywords

polysemy; lexical semantics; ambiguity; underspecified representation; eye movements; psycholinguistics

Polysemy is one of the central issues in the psychology of word meaning. Most common content words do not have a single, simple meaning but instead pick out a number of related meanings, or *senses*. Two important questions are how those senses are represented in the lexicon and how they are processed during language comprehension. In this paper, we focus on apparently conflicting findings regarding the representation of polysemous senses. On the one hand, there is evidence showing that there is little semantic overlap between senses, supporting the view that senses of a polysemous word must then be represented separately (Klein & Murphy, 2001, 2002), like the completely distinct meanings of homonyms. On the

Please address all correspondence to: Stephani Foraker, Department of Psychology, SUNY College at Buffalo, 1300 Elmwood Ave., Buffalo, NY 14222, forakesm@buffalostate.edu, phone: 917-622-2180, fax: 716-878-6228.

other hand, studies investigating sentence comprehension have concluded that polysemous words have an underspecified meaning that encompasses its different senses (Frazier & Rayner, 1990; Frisson & Pickering, 1999; Pickering & Frisson, 2001). We present three experiments that contrast these views of sense representation, using a sentence processing approach.

Sense Relatedness and Sense Similarity

One critical aspect of polysemy is that the different senses of a word can be closely *related* but not very *similar* to each other. For example, consider the word *church* used to refer to a building sense, as in *The church burned down*, versus an organization sense, as in *The church has lost many members*. These senses are clearly closely related, as the organization built the building and carries out its activities there. Nonetheless, buildings are not conceptually *similar* to organizations: One of them has bricks, mortar, and electrical wires, is a certain height, and has a color and weight; the other is a cultural and legal organization that people join, with a set of beliefs and practices. None of these things can be said about the other—the building is not a legal entity that people join, and the organization is not made of bricks and doesn't have a color. Klein and Murphy (2002) found that people did not generally sort together the different senses of the words in a categorization task, nor did they induce a novel property of one sense to the other sense (e.g., they judged that a new property of a church building would be unlikely to be found in the organization). By usual measures of conceptual coherence, the different senses of polysemous words did not form coherent categories. Therefore, although there is often a very obvious connection between different senses of a polysemous word, the senses may nonetheless have few semantic features in common.

This difference has a functional consequence for the processing domain as well, namely that switching from one sense of a polysemous word to a different one creates a processing cost in a semantic judgment task (Klein & Murphy, 2001; Pylkkänen, Llinas, & Murphy, 2006). In contrast, using the same word twice in the same sense facilitates processing. These results suggest that the amount of meaning that is common across different senses of a word cannot be very substantial, or else facilitation would have been found rather than interference when the same word was used in different ways. These conclusions are consistent with a number of linguistic analyses of polysemy, which have pointed out the large differences in the meanings of different senses of various words (e.g., Cruse, 1986; Rice, 1992).

Within the psycholinguistic literature, most research on polysemy has focused on how polysemous words are processed, often comparing them to homonyms (like *bank* or *calf*), which have completely unrelated and incompatible meanings. When readers encounter homonyms with two frequent meanings, they immediately activate both meanings (Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979). However, if one meaning is significantly more frequent than the other, then that meaning tends to be more activated, and the less frequent one less so (Duffy, Morris, & Rayner, 1988). Some authors have argued that this quick resolution of homonyms may be necessary for comprehension, as the very different meanings are "semantically incompatible" (Frazier, 1999, p. 39). In a neutral context where one cannot be sure whether the speaker is referring to a financial institution or the side of a river (for *bank*), the different meanings lead to completely different interpretations of the subsequent discourse. Quick resolution may be necessary because maintaining the ambiguity of these very different possibilities may be costly. Thus, the fast selection of the dominant meaning of a homonym, even without contextual support, may be an effective strategy for dealing with the ambiguity.

It is natural to compare this strategy to how people might deal with polysemy, which is also a form of ambiguity. Frazier and Rayner (1990), Frazier (1999), and Pickering and Frisson (2001) have argued that polysemy may show a rather different pattern of results, because the meanings of polysemous words are related in a principled way (see Frisson, 2009, for a review). In particular, they have argued that one might safely postpone a fully committed decision about which sense of a polysemous word was intended, because of their relatedness. One does not need to make an early commitment to what the speaker meant by *church*, because the two senses are more compatible than the unrelated meanings of *bank* are. Furthermore, they argue that in a neutral context, a reader activates an underspecified, abstract semantic representation of the polysemous word that encompasses or underlies both senses (Frazier & Rayner's partial specification hypothesis, 1990). Following the polysemous word, subsequent information available in the sentence will guide the fuller interpretation of a particular sense. And indeed, these studies report results for reading sentences with polysemous words, following neutral context, that were different from results concerning homonyms.

In the case of Frazier and Rayner (1990), the results were qualitatively different for homonyms and polysemous nouns. Readers seemed to pick the dominant meaning of a homonym when there was no constraining context, but with polysemous words, readers showed no preference for the dominant over the subordinate sense. Pickering and Frisson (2001), though, found that when there was no constraining context preceding an ambiguous verb, both homonymous verbs with multiple meanings (*rule*) and polysemous ones with multiple senses (*disarm*) allowed readers to compute the dominant sense more easily, but such frequency effects occurred later for polysemous verbs. Both sets of authors have argued that these differences reveal that people are taking advantage of the related senses of polysemous words, which do not require an immediate commitment, as the unrelated meanings of homonyms do.

Core Meaning

The argument about the difference between homonyms and polysemes depends on some assumptions about the representations of these words and how they are processed. Polysemous words are often said to share a *core meaning* that is fairly constant across different senses (e.g., Caramazza & Grober, 1976; Ruhl, 1989). If that is the case, then when encountering a polysemous word with no biasing context, one could simply retrieve the core meaning and have some idea of what the speaker is talking about. Later information would add more features to this core, resulting in the specific, intended sense. As homonyms do not have this shared core meaning, such a strategy would not be possible for them.

However, this notion of a core meaning for polysemous words has been criticized within the linguistic literature (Cruse, 1986, p. 71–4; Lakoff, 1987; Murphy, 2007; Rice, 1992; Zgusta, 1971, p. 66), and as we pointed out earlier, Klein and Murphy's (2001, 2002) experiments suggest that many polysemous senses have very few features in common. If people read the word *church* in a neutral context, they cannot evoke a core meaning encompassing both the building and the religious organization, because the two share few, if any, common features. Furthermore, Klein and Murphy (2001) found results for comprehending polysemous words that were only slightly different from results with homonyms. Thus, if one interprets "core meaning" to be the components of a word's meaning that are common across all its senses, then this hypothesis about how polysemous words are represented is problematic. The fact that senses can be formally related (e.g., *church* as an institution and as a building housing that institution) does not necessarily entail that they have many semantic features in common.

The notion that readers do not commit to a single sense of a polysemous word may still be possible, however, if one does not claim that a core meaning is accessed. Perhaps readers simply do not select any sense of a polysemous word in a neutral context but instead wait for context to indicate the appropriate sense. Although it is not clear why they should do this for polysemous words and not for homonyms, the reason may be due to the relation between the senses, which allows either sense to be accessed quickly in a way that the unrelated meanings of homonyms do not (see General Discussion).

Another possibility is that polysemous words differ in this respect. Perhaps words that have *similar* senses do allow a core meaning to be accessed, whereas words with more distinct senses do not. Polysemy is a broad phenomenon, and some senses that theorists have identified for a word may be fairly similar to one another. For example, Klepousniotou, Titone, and Romero (2008) argued that the senses in object/substance polysemy (e.g., *rabbit* referring to the animal vs. meat) overlap semantically, and that a word like *rabbit* might evoke a core meaning when it is first accessed. Another frequent form of polysemy (Nunberg, 1979) is when a noun refers to the entire class vs. specific entities (e.g., *the dog has been domesticated for centuries* vs. *the dog lay down*). Although these are ontologically very different entities, the features of the class of dogs seem mostly true of individual dogs. Klepousniotou et al. (2008) compared words that had senses with high, medium, or low overlap, and found evidence that low overlap between senses increased the switching cost from one sense to another. They suggested that Klein and Murphy's (2001) finding of a large cost was due to their use of polysemous words that generally had dissimilar senses.

The Present Experiments

Previous studies have led to apparently conflicting conclusions. On the one hand, there is evidence that polysemous senses are rather different from one another (Klein & Murphy, 2002), and corresponding processing effects showing that switching from one sense to another has a significant cost (Klein & Murphy, 2001). The processing effects have been argued to be most consistent with a view in which different senses are explicitly represented in the lexicon and compete with one another at access (Klein & Murphy, 2001; Pylkkänen et al., 2006). On the other hand, some studies have found that with a non-biasing or neutral preceding context, different senses of a polysemous word are reached with similar ease (Frazier & Rayner, 1990), or at least, the amount of difficulty is later than that encountered for different meanings of a homonym (Pickering & Frisson, 2001). This has been taken to suggest that a core or underspecified meaning of a polysemous word can simultaneously represent different senses. However, that in turn suggests that the different senses of a single word are functionally alike, perhaps sharing considerable meaning.

It is in fact difficult to determine whether there is a real conflict between these two sets of studies for three reasons. First, they were designed with different goals. Murphy and colleagues focused on the representation of polysemy, whereas the reading studies focused on processing issues, such as the time course of sense activation. These different goals were pursued using different methods. For example, Klein and Murphy (2001) examined priming in a sensicality task in which people had to explicitly judge the meaningfulness of a phrase containing the target word (e.g., *brick church*). Frisson (2009) suggested that explicitly judging sensicality may emphasize the distinctive meaning of each sense. In contrast, Frazier and Rayner (1990), Frisson and Pickering (1999), and Pickering and Frisson (2001) used eyetracking of sentence reading, providing only one instance of a sense at a time. Also, a focus of some of the reading studies was meaning dominance effects, which Klein and Murphy did not examine. Although differing conclusions have been reached, it is unclear whether this is a true empirical disagreement or simply differences due to different tasks or experimental designs.

Secondly, as mentioned, there are different forms of polysemy, which likely have different properties. For example, Frazier and Rayner (1990) used only concrete versus abstract senses of a noun, which were productive metonymies (e.g., *dinner* as an event or food, *book* as an object or publication). Frisson and Pickering (1999) investigated literal versus metonymic senses of a noun, such as place/institution (e.g., *convent*) and place/event (*Vietnam*), and Pickering and Frisson (2001) examined literal versus metaphorical senses of verbs (e.g., *disarmed*). Murphy and colleagues, on the other hand, have used diverse forms of polysemy in their experiments. Thus, conflicting findings may also be accounted for to the extent that the conflicting studies have used materials containing different forms of polysemy. It is unfortunate, then, that all these experiments differ both in materials and the tasks used.

Finally, the conclusion of significant shared meaning of different polysemous senses was supported in the reading studies of nouns by null results—absence of processing cost for the subordinate sense following neutral context (Frazier & Rayner, 1990; Frisson & Pickering, 1999). It is possible that the disambiguating contexts or sense frequency manipulations were not strong or unambiguous enough to produce an effect in reading times.

To help resolve these differences, the present studies used the polysemous words from Klein and Murphy (2001), which provided evidence for separate sense representations, and switched to a sentence reading task like those that have provided evidence for core meaning or underspecified representations. We examined two issues investigated in the earlier reading studies: (1) how context influences the interpretation of polysemous words and (2) whether there is a preference for a dominant sense in the absence of a constraining context, a *sense frequency* effect. If our results are similar to theirs, then we can conclude that the apparent differences in past work are likely due to task variables (reading vs. sensicality judgments); if our results are different from theirs, even using reading as a dependent variable, we can conclude that the differences are likely due to stimulus differences. The results will speak both to questions of how the senses of polysemous words are represented in the mental lexicon and how these senses are constructed or accessed during reading.

Experiment 1

Experiment 1 used the polysemous words from Klein and Murphy (2001). We constructed sentence contexts that would prime one of the senses or else be neutral. Consider the sentences for *cotton* shown in Table 1. The contexts used a subject noun phrase that was consistent with one sense more than the other. Presumably fashion designers are more interested in fabric than they are in plants, whereas farmers have the reverse preference.

Following the context sentence, subjects read one of two target sentences, consistent with either the dominant or subordinate sense. These sentences were identical except for the initial noun phrase. When the context picked out one sense of the polysemous word in the first sentence, target sentences consistent with that sense should be easier to read. We also predicted that in these consistent cases, the subordinate and dominant senses would be equally easy to arrive at, since the biasing contexts were effective in supporting their given sense. For inconsistent target senses, though, we predicted that sense frequency would play a role, as re-access or reanalysis would be necessary. The more frequent, dominant sense should be easier to retrieve or compute upon reanalysis, while the less frequent, subordinate sense should be more difficult. This would suggest that different senses are represented separately in the lexicon, with the most frequent ones having stronger representations.

The critical question was what happens with neutral context, when there is no priming of a particular sense of the polysemous word. According to an underspecification account (Frazier & Rayner, 1990; Pickering & Frisson, 2001), readers activate a core meaning that is equally consistent with either sense, or make no commitment to either sense until further information is made available to support “homing in” on the appropriate one. Under these circumstances, there would be no preference for the dominant over the subordinate target sentence. Alternatively, if senses are stored and represented separately, their retrieval should be sensitive to frequency of occurrence. Under this assumption, if people select the more frequent sense of a polysemous word following a neutral context, they should then be faster at reading the target sentence when it instantiates the dominant sense rather than the subordinate sense.

Method

Participants—Participants in this and the following experiments were monolingual American-English speakers with no history of reading difficulties. They were students at New York University, receiving course credit or pay for participating. Thirty-six participated in Experiment 1.

Materials—From Klein and Murphy (2001), 24 polysemous nouns were selected, with the addition of *school*, to total 25. These words were originally derived from published lists of polysemous words and then checked against a dictionary, which listed both of the tested senses as senses of the same word, with a shared derivation (unlike homonyms; see Rodd, Gaskell, & Marslen-Wilson, 2002, for similar methodology). We developed target sentences instantiating the two different senses of each word in cycles of pretesting to ensure that the senses intended were the ones arrived at by readers.

The stimuli consisted of sentence pairs. The first sentence provided a context for the polysemous word, which ended the sentence. Context was biased towards the dominant or the subordinate sense or else was neutral. The verb of this sentence was held constant and was not clearly predictive of either sense. Thus, bias towards a sense was accomplished via the sentence subject (e.g., *fashion designers* or *farm owners* in the case of *cotton*). The subject of the neutral context was a pronoun. The second sentence, the target sentence, began with a noun phrase that was closely associated with either the dominant or subordinate target sense, thereby disambiguating the polysemous word. This produced six conditions in a 3 × 2 design: dominant, subordinate, or neutral context crossed with the dominant or subordinate target sense. Across the 25 items, the disambiguating subject noun in the dominant and subordinate target sentences was equated as much as possible for length (*M*s: dominant = 7.6 letters, subordinate = 7.0 letters), and lexical frequency (*M*s: dominant = 34 per million, subordinate = 31; *medians*: dominant = 15, subordinate = 15; Francis & Kuçera, 1982). The full set of materials appears in an online supplementary appendix.

Pretesting for sense bias: In order to identify the dominant and subordinate senses of the polysemous word, we first administered a sentence completion task. Thirty participants were given the neutral context sentence followed by a pronoun referring to the polysemous word, which started the fragment to be completed (*They discussed the cotton. It* ____). Participants were instructed to complete the sentence with whatever came to mind first. Three raters coded the completions for the two senses used in Klein and Murphy (2001). Instances were discarded when neither sense was used or when the sentence was consistent with either sense. The dominance scores for each item were highly correlated across raters ($r = .93$–$.97$). The sense with greater probability of occurring following a neutral context was labeled “dominant” while that with the lesser probability was labeled “subordinate.” One item, *school*, was rated separately: 26 participants wrote completions, and two of the previously mentioned coders rated them, again with high reliability ($r = .92$). The average sense frequencies and sense glosses are presented in Table 2, showing a range of sense frequency.

Pretesting to confirm the target sentence sense: Once the dominant and subordinate senses were ascertained, a separate group of participants was asked to choose what the polysemous word meant in the neutral context conditions. In two groups, 24 participants were presented with a two-alternative forced-choice for each item. For example, participants first read *They discussed the cotton. The [fabric/crop] was not what they had been hoping for.* Half saw *fabric* in the sentence, and half saw *crop*. They were then asked: "What does 'cotton' mean here? Type of thread or cloth OR plant containing a soft white substance." The two versions of the second sentence, each consistent with only one sense of the word, were changed and retested as needed, until at least 80% of participants chose the meaning intended.

Pretesting the understanding of target sentences: The final pretest was to ensure that any differences found in the reading time experiment would not be due to simply preferring one of the target sentences over the other, perhaps due to plausibility or familiarity. Sixteen participants were divided into two counterbalanced groups and presented with the target sentence on a computer screen, embedded randomly within an equal number of unacceptable sentences. For example, *The crop was not what they had been hoping for* should be judged acceptable, while *The lamp walked down the street* should not. Using E-prime software (Schneider, Eschman, & Zuccolotto, 2002), we recorded whole sentence reading times and judgments of whether a sentence was acceptable. Participants accepted the dominant and subordinate target sentences equally often (dominant = .95, subordinate = .93, $t(15) = 1.08$, $p > .29$) and rejected unacceptable ones at a high rate (.93). More importantly, there were no reading time differences between the target sentences (M dominant = 2305 ms, subordinate = 2377 ms, $t(15) = 1.20$, $p > .24$).

Procedure—Participants sat in front of a PC monitor, viewed one sentence at a time, and were instructed to read normally for comprehension. Once they had finished reading and fully understood a sentence, they pressed the space bar to reveal the next one. They completed eight practice trials first and were aware that the sentences were to be read as pairs. Whole-sentence reading times were recorded. The experiment took about half an hour.

The 25 items and 6 conditions were divided among 6 counterbalanced lists to avoid any repetition of an item for a given participant. Thus, each participant encountered each critical word once and served in all conditions, across items. The experimental items were embedded in an equal number of filler items of similar structure that avoided confusing polysemy as much as possible (e.g., *He hung the curtains. The rod was difficult to reach without a ladder.*). The fillers were intended to reduce the chance of participants noticing lexical ambiguity in the materials.

Data analysis—Reading times shorter than 50 ms were excluded, which constituted less than 2% of the data. Reported here are analyses on reading times with no further exclusions. (Analyses of the data trimmed by 3 *SD*s around subject or item means produced similar results to those reported here.) All analyses were performed with participants (F_1) and with items (F_2) as a random factor. F_1 analyses include the between-participants counterbalancing term.

Results and Discussion

For the first sentence, which ended with the polysemous word, we found a significant main effect of context, since the shorter neutral context sentences, beginning with a pronoun (M = 1491 ms, SE = 79), were read more quickly than the longer biasing ones, which began with a full noun phrase, $F_1(2, 60) = 67.59$, $p < .001$; $F_2(2, 48) = 21.48$, $p < .001$. A further comparison of the two biasing context sentences showed that they did not differ from one another, *F*s < 1, indicating that the dominant and subordinate contexts were equally easy to process (dominant *M* = 1882 ms, *SE* = 82, subordinate *M* = 1904, *SE* = 88).

For the critical target sentence, which began with a noun phrase instantiating either the dominant or subordinate sense, the full 3 (context) × 2 (sense) ANOVA showed a significant interaction, $F_1(2, 60) = 23.08$, $p < .001$; $F_2(2, 48) = 12.73$, $p < .001$, illustrated in Figure 1. We now present the findings relevant to the experimental questions in more detail.

Effects of biasing context—When the disambiguating target sentence was consistent with the sense primed by the preceding context, processing was easier than when it was inconsistent. Following the dominant context, the dominant sense (*M* = 2178 ms, *SE* = 129 ms) was read more quickly than the subordinate sense (*M* = 3083, *SE* = 189), and following the subordinate context, the subordinate sense (*M* = 2343, *SE* = 158) was read more quickly than the dominant sense (*M* = 2777, *SE* = 174), resulting in a significant 2 × 2 cross-over interaction of context and target sense, $F_1(1, 30) = 36.82$, $p < .001$; $F_2(1, 24) = 19.52$, $p < .001$. Paired comparisons showed no difference between the two consistent conditions, such that subordinate and dominant targets were equally easy to process when preceded by their appropriate contexts, *p*s > .10. On the other hand, recovering from an inconsistent context differed depending on sense frequency, consistent with our predictions. Reanalyzing to reach the subordinate sense given a dominant context was harder than reaching the dominant sense given a subordinate context, $t_1(35) = 2.10$, $p < .05$, $t_2(24) = 1.29$, $p = .21$. The effects of consistent versus inconsistent biasing context confirm that our materials produced a particular interpretation of the polysemous word and, most importantly, that reading time of the target sentence was influenced by readers' interpretation of the earlier context sentence.

Neutral context—Following a neutral context the dominant sense (*M* = 2213, *SE* = 115) was read significantly faster than the subordinate sense (*M* = 2669, *SE* = 160), $t_1(35) = 4.00$, $p < .001$, $t_2(24) = 2.60$, $p < .02$. This does not support an underspecified representation of a polysemous word, which predicts that the two senses should be filled out with equal ease.

The underspecification hypothesis further predicts that it should take longer to process a sense following neutral context than following consistent context. Since the sense would not be fully specified, there should be some cost to computing or filling out the correct meaning upon integration with the target sentence (as found by Pickering & Frisson, 2001). Because we predicted that readers would derive the dominant sense in the neutral context, we can compare this condition to the same sense with contextual support (dominant context-dominant sense). In fact, there was no difference between these two conditions, *p*s > .50, illustrated in Figure 1, suggesting that even in the neutral condition readers had activated the dominant sense quite strongly.

Summary—These results are most consistent with the notion that when readers encounter a neutral sentence such as *They discussed the cotton*, they generally interpret *cotton* as indicating the dominant sense, cloth made of cotton, rather than remaining noncommittal between that sense and other senses (e.g., the crop). Following neutral context, the target sentence was read reliably faster when it picked out the dominant sense. Furthermore, adding preceding context that biased the word towards the dominant sense did not speed reading of the dominant sense target sentence, again suggesting that readers had already accessed the dominant sense, as a default meaning. In sum, these results are most consistent with the claim that the dominant sense is usually accessed, rather than with the notion that an underspecified meaning, which is equally compatible with both senses, is evoked.

Experiment 2

In Experiment 1, we constructed the materials with the polysemous word and disambiguating noun in different sentences. The major reason for this was to keep the words and their subsequent disambiguations separate, reducing potential spillover. However, this aspect differs from previous reading studies finding evidence supporting an underspecified representation, which employed one-sentence materials presented on one display (Frazier & Rayner, 1990; Frisson & Pickering, 1999; Pickering & Frisson, 2001). It is possible that in Experiment 1 the polysemous words were in fact underspecified when first encountered but that this representation was not carried past a sentence boundary (albeit one that occurred immediately after the polyseme). In Experiments 2 and 3, then, we incorporated the context, polysemous word, and disambiguating region into a single sentence to test whether interpreting a polysemous word within a sentence is the same as processing across sentences. Experiment 2 investigated one-sentence materials using self-paced reading, while Experiment 3 used eye-tracking. Although we now test single-sentence materials, it should be kept in mind that in much text, a polysemous word will not be immediately disambiguated by context and the Experiment 1 materials provide a model for that situation.

We predicted the same pattern of processing as for Experiment 1, focusing on effects in the disambiguating region. First, consistent context-target sentences should be easier to process than inconsistent ones (and the two consistent conditions should be equally easy). Second, sense frequency should have measureable effects for the inconsistent conditions, such that misleading dominant context should be more difficult to recover from than misleading subordinate context. Third, following neutral context, the dominant sense should be easier to process than the subordinate sense, consistent with a default commitment to the dominant sense.

Method

Participants—Thirty-six people from the same population participated.

Materials—We employed the same six conditions (see Table 1), polysemous words, and senses as in Experiment 1. The stimuli consisted of one sentence, with context preceding the polysemous word (e.g., *cotton*), and a disambiguating target phrase following it (e.g., *after the fabric ripped*).

The words in the disambiguating sense region were controlled for length and lexical frequency between conditions. For the word that was first or central to the target phrase (e.g., *crop* and *fabric* in the above example), dominant sense targets had a mean length of 9.2 letters, a mean frequency of 87 per million and a median frequency of 20 per million, while subordinate-sense noun phrases had a mean length of 9.1 letters, a mean frequency of 85, and a median frequency of 41. The full set of materials appears in the online supplement.

The materials were again pretested to ensure that the senses intended by the disambiguating sense phrase were the ones understood by readers. Twenty-four participants, divided into two groups, judged which sense was indicated by the neutral contexts (e.g., "*They discussed the cotton after the fabric ripped a second time.* What does 'cotton' mean here? Type of thread or cloth OR Plant containing a soft white substance"). As in Experiment 1, the disambiguating sense phrase (*after the fabric ripped/crop failed*) was changed and retested as needed until at least 80% of participants chose the intended meanings.

Procedure—Participants read each sentence for understanding on a PC, using a phrase-by-phrase, noncumulative self-paced moving window procedure, controlled by E-prime. The reading time for each phrase was recorded. Untimed comprehension questions (yes-no) followed half of the trials, with accuracy feedback provided.

Sentences were divided into five regions: Context Noun Phrase (NP), main Verb, Polysemous NP, Disambiguating region, and Spillover region, indicated by carats (^) in this example: *The fashion designers* ^ *discussed* ^ *the cotton* ^ *after the fabric ripped* ^ *a second time*. Following eight practice trials, the stimuli were presented randomly among an equal number of fillers that avoided obvious or confusing polysemy (i.e., one-sentence versions of the Experiment 1 fillers). The whole procedure took about half an hour.

Data analysis—Reading times shorter than 50 ms were excluded, which constituted less than 2% of the data. (Again, analyses of the data trimmed by 3 *SD*s around subject or item means produced similar results to those reported here.) Participants answered the comprehension questions with a high degree of accuracy ($M = .96$).

Results and Discussion

Table 3 presents the reading times for each region in Experiment 2, showing the same pattern of effects as Experiment 1, although not as strongly. The only significant effect found in the Context NP region was the uninteresting effect of context, due to faster reading of the shorter neutral contexts, $F_1(2, 60) = 49.35$, $p < .001$, $F_2(2, 48) = 30.84$, $p < .001$. This pattern persisted into the following Verb region, $F_1(2, 60) = 7.69$, $p = .001$, $F_2(2, 48) = 4.12$, $p = .02$. On the polysemous word itself there were no significant differences, in the full design or more specific comparisons, *p*s > .47, as in Experiment 1.

Table 4 presents the inferential statistical tests for the Disambiguating and Spillover regions. (Main effects for context and sense factors are reported for completeness, but do not bear directly on our hypotheses.) In the Disambiguating target sense region, the full ANOVA showed a significant interaction, which persisted into the Spillover region, although not significant by items. We therefore turn to our more specific hypotheses.

Effects of biasing context—As in Experiment 1, the biasing context was effective. Consistent conditions were read faster than the inconsistent ones, as shown in Table 3. Table 4 reveals that the 2 × 2 Context by Sense interaction was significant by participants and items at the Disambiguating phrase, and significant by participants at the Spillover phrase. Further comparisons for the two consistent conditions showed no difference in reading time for the Disambiguating or Spillover phrases, *p*s > .38. These results again show evidence that sense frequency affected readers' interpretations in the inconsistent conditions. In the Disambiguating and Spillover regions readers found it more difficult to reanalyze the meaning of a polysemous noun following dominant biasing context compared to reanalysis after subordinate biasing context. Although not particularly surprising, this result is important in showing that the contexts did bias interpretation and then influence subsequent sentence interpretation in the intended way.

Neutral context—As shown in Table 4, following neutral context, the dominant sense continuations were read marginally faster than the subordinate sense continuations in the Disambiguating and Spillover regions (not significant by items). Because the effect of dominance might have been spread across the Disambiguating and Spillover regions, we performed analyses of those two regions combined, which revealed a strong effect by participants, $t(35) = 3.14$, $p < .01$, but still unreliable by items, $t(24) = 1.63$. Pursuing the dominant sense advantage further, we found that the dominant sense was read as quickly following neutral context as when it followed a consistent, dominant context (in fact, it was numerically faster, Table 2) in the Disambiguating and Spillover phrases, *p*s > .35. Thus, as in Experiment 1, there is no additional benefit to having consistent preceding context when one is accessing the dominant sense, contrary to an underspecification proposal.

Sense frequency and similarity—Recall that we used Klein and Murphy's (2001, 2002) items specifically to discover why their findings conflicted with those supporting an underspecified representation (Frazier & Rayner, 1990; Pickering & Frisson, 2001). Klein and Murphy's conclusion that related senses are separately represented in the lexicon predicts that senses should show frequency effects, such that a more frequent sense should be more available, leading to a faster reading time than a less frequent sense. However, Klein and Murphy did not select their items so that each one would have a clearly dominant sense. As seen in Table 2, not all of the items show a sizeable difference between the frequency of the dominant and subordinate senses. For example, *coat* occurred in the dominant sense 91% of the time and the subordinate sense 1% of the time, whereas *sheets* was used 27% and 19% respectively (the other responses being too vague to assign to either sense or indicating a different sense). Some show a pattern more similar to balanced frequency (*school, filling, cotton, chicken, shower, fortune, sheets*), which would weaken any dominance effect (Duffy et al., 1988). If our predictions are correct, then the dominance effect in neutral contexts should increase as the strength of the dominant sense increases compared to the subordinate sense. We therefore calculated a dominance score for each item by taking the ratio of the dominant sense frequency to the subordinate sense frequency (based on Table 2).

We also investigated whether the dominance effect depended on the similarity of the two senses, as proposed by Klepousniotou et al. (2008). We asked 16 participants to rate the similarity of the two senses on a scale from 1, completely different, to 7, almost identical. Participants read the two consistent conditions (The farm owners discussed the COTTON after the crop failed once again. The fashion designers discussed the COTTON after the fabric ripped once again.) and then rated how similar the two meanings of the capitalized word were (i.e., share common properties, like shape, composition, or use). As shown in Table 2, the mean similarity ratings ranged from 1.37 to 6.56 (M = 3.90, SD = 1.63). The ratings for the 25 polysemes were on average higher than those for 25 homonyms (M = 1.49, SD = 0.43) and much lower than those for 20 synonyms (M = 6.40, SD = 0.43) that were included in the norming questionnaire.

We submitted the reading time data to a linear mixed-effects regression model. The fixed factor predictor terms were the sense completion (dominant or subordinate), dominance ratio, similarity rating, and region length, which were first centered. (Models that included lexical frequency residualized on region length as a fixed factor showed similar results so are not reported here for ease of exposition.) Analyses were conducted using the lme4 (Bates, Maechler, & Bolker, 2011) and languageR libraries (Baayen, 2011) for the R statistics program (The R Foundation for Statistical Computing, 2012). Participants and items were included as cross-classified random factors in the model, and all interactions between sense completion, dominance ratio, and similarity rating were included in initial models (as well as the counterbalancing list term for subjects). The three fixed factors of theoretical interest were not highly correlated, reducing potential co-linearity concerns. Reading times were trimmed by 3 SDs (by subjects and items) as part of the fitting routine. The simplified models reported here yield the best-fit maximal random effects structure justified by model comparison procedures (Baayen, Davidson, & Bates, 2008; Jaeger, 2009), while retaining the factors of theoretical interest to compare effects across regions and experiments (as shown in Tables 5 and 8). If the absolute t-value for a fixed factor was over 2, the effect of the fixed factor was considered significant at = .05 (Gelman & Hill, 2007), and 1.80 was considered marginal. We were not able to supply more specific p-values, since

Markov chain Monte Carlo sampling for models with random slopes and intercepts has not yet been implemented in this software.

For the neutral context conditions, we expected to find an interaction of dominance proportion and sense, indicating that the stronger the dominant sense was, the greater its reading time advantage would be. On the Polysemous NP, the only significant effect found was of region length, $t = 5.49$. As shown in Table 5, on the following Disambiguating sense phrase, there was a significant length effect, with longer regions producing longer reading times. Consistent with our prediction, the dominance score interacted marginally with sense completion, indicating that the greater the frequency spread between the dominant and subordinate senses, the greater the difference between the reading times for the two senses. There was also a marginal effect of similarity, with shorter times for greater similarity. Similarity did not interact with the dominance score or sense completion. On the Spillover, the significant effect for sense completion indicated that the subordinate sense completion took longer to read than the dominant one, and there was again a significant effect of similarity.

Thus, consistent with separate sense representations, we found that the reading time advantage for the dominant sense on the disambiguating phrase reflected how dominant that sense was. Readers were biased to interpret a polysemous word in its more frequent sense, and this effect occurred in addition to any contribution of sense similarity. The absence of any interaction involving sense similarity suggests that the sense frequency effects do not differ for words with less similar vs. overlapping senses (e.g., *trunk* vs. *book*), although the latter are read faster in general.

Summary—On the polysemous word itself, we found no effects, consistent with Experiment 1. On the subsequent regions, we found that inconsistency between context and target senses significantly slowed reading compared to consistent use and that sense frequency affected the ease of recovering from inconsistent context.

After a neutral, nonbiasing context there was an advantage for choosing the dominant sense of the word, although it was statistically not as strong as that found in Experiment 1. However, this was likely due to variations in the degree of sense dominance across items, which explained a significant proportion of reading time in the neutral context conditions, separately from sense similarity. Furthermore, when accessing the dominant sense of a polysemous word, prior context consistent with the dominant sense provided no processing advantage over neutral context. This finding is contrary to a core or underspecification account. These results are most consistent with theories in which senses are separately represented.

Experiment 3

Experiment 3 examined the same polysemous words as the previous two experiments using eye-tracking measures of reading. The first goal of this experiment was to assess the finer-grained time course of comprehending polysemous nouns, focusing on the source of the sense dominance effects we found in the previous two experiments. In Experiment 2, the dominance effect in the neutral context conditions was not fully reliable, and examining the multiple dependent measures of processing difficulty that eye-tracking affords can provide a more detailed picture of when separate senses may be accessed or fully committed to in the one-sentence materials.

The second goal of Experiment 3 was to provide a more similar task for comparing our dominance effects with previous studies that did not find effects of sense frequency (Frazier

& Rayner, 1990; Frisson & Pickering, 1999). Experiments 1 and 2 provided evidence that readers select an individual sense representation for the kinds of polysemous words employed by Klein and Murphy (2001), even when read in sentences rather than appearing as modified nouns in a semantic judgment task. However, Experiments 1 and 2 used self-paced reading, while previous experiments that have not found effects of sense frequency following neutral context (Frazier & Rayner, 1990) used eye-tracking during reading. We predicted that biasing context conditions would show the same interaction between context and disambiguating sense that we found in the previous two experiments and that there would be dominance effects in neutral contexts.

Researchers have suggested that effects in early measures, such as how long a region is first fixated, reflect lexical access and are reliably influenced by lexical frequency, meaning frequency or degree of ambiguity, and previous context. Later measures, such as how many times or for how long a region is refixated, tend to reflect different aspects of processing, such as post-lexical integration, checking processes, or reanalysis (for reviews see Pickering, Frisson, McElree, & Traxler, 2004; Rayner, 1998). The dominance effects we are predicting may happen relatively quickly (e.g., First Pass reading), suggesting immediate access to the dominant sense. On the other hand, dominance effects could occur later (e.g., Second Pass reading), suggesting that the senses take time to be retrieved or integrated into the sentence representation. Consider that disambiguating NPs of the sort we used do not strictly determine the polysemous word's interpretation. For example, after a reference to *cotton*, the phrase *The fabric* probably indicates that cloth is being referred to, but it could also be the beginning of a reference to something else (*They planted the cotton. The fabric of their shirts was soaked with sweat*). It is only after reading the entire sentence that disambiguation can be certain.

Previous eye-tracking experiments examining polysemous nouns have not included our exact conditions. Frazier and Rayner (1990) compared reading time patterns for dominant vs. subordinate interpretations of a polysemous noun. When biasing context preceded the polyseme, they found a weak indication that the unpreferred sense was more difficult to process than the preferred sense. However, the materials provided no way to assess which sense a reader had reached. For example, the sentences *Lying in the rain, the newspaper was destroyed*, and *Managing advertising so poorly, the newspaper was destroyed*, do not have material after "newspaper" similar in function to our Disambiguating region, making comparison with our biased conditions difficult. Comparing the experiments is feasible in neutral contexts, which both had disambiguating information after the polysemous noun (*Unfortunately, the newspaper was destroyed, lying in the rain*/*managing advertising so poorly*). In those cases, Frazier and Rayner (1990) found no differences between the two senses, for any region or measure.

In Frisson and Pickering (1999), the conditions relevant to our study used context preceding a polysemous noun that biased either a familiar literal sense or a familiar metaphorical sense of the noun. They found that sense frequency did not correlate with differences in reading time or eye-movement regressions when comparing literal versus metaphorical senses. They also performed a secondary analysis on high-frequency vs. low-frequency senses, but they did not test a neutral context condition, which is where frequency effects should have a strong effect.

Method

Participants—Fifty-four people from the same population as Experiments 1 and 2 participated. All had normal or corrected vision (contacts).

Materials—We used the same design and 25 items as Experiment 2, with minor changes to the structure of the sentences, but no content changes to the biasing context, polysemous word, or disambiguating regions. The main change was to lengthen the region between the polysemous word and the disambiguating sense region, to increase the distance and time available for the potential filling out of an underspecified sense (Pickering & Frisson, 2001). For example, *The caterer tested the filling in the pie before leaving* became *The caterer tested the filling that was in the pie before leaving.* The "between" region was 2–4 words long ($M = 2.9$, $SD = .9$; number of characters $M = 13.0$, $SD = 4.0$). We also made minor revisions to the critical disambiguating regions for some items, to equate them more precisely. For example, *chicken that was baking in the oven/squawking in the coop*, became *chicken that was in the oven/coop*. The disambiguating region was either a noun or a two-word phrase, which was equated within a particular item for number of words and for character length as closely as possible. Because we did not change the root words at all, the lexical frequency of the disambiguating region remained equated for the two senses, as in Experiment 2. The full set of materials appears in the online supplement.

The 25 items in the 6 conditions were distributed among 8 counterbalanced lists, with 6 lists containing 19 items and 2 containing 18 items (lists were uneven because materials were combined with an unrelated 8-condition experiment). Thus, each participant saw three or four instances of each condition. The experimental items constituted 10% of the total trials ($N = 180$) in a session. The fillers did not include polysemous words or homonyms with intentionally confusing uses, and were all one sentence. Example filler sentences included *The mother answered the phone after the ringing woke her; Timothy stamped his feet and muttered, but nobody knew what, because he is so quiet; The bully who pushed the man into the table this afternoon left in a hurry*. A yes-no comprehension question followed half of the trials, which asked about the context noun, action or verb, polysemous noun, and disambiguating region equally often.

Procedure—We collected eye-tracking data with a SensoriMotor Instruments EyeLink II head-mounted eye-tracker apparatus and software, interfaced with EyeTrack presentation software (Stracuzzi & Kinsey, 2006). The eyetracker records eye movements and fixations binocularly every 4 ms. Participants used a chin rest at a distance such that 1° of visual angle subtended 3 characters. Calibration and validation of the recording apparatus were performed before beginning the experiment; drift correction was applied before every trial. Participants were randomly assigned to a list, and the 180 trials appeared in a unique random order for each participant.

Participants were instructed to read for comprehension in a natural way. The sentence appeared all at once on one line and remained on the screen until the participant finished reading, signaled with a mouse click. Comprehension questions appeared on a subsequent screen and were also answered yes/no using the mouse. Participants had five practice trials before beginning the experiment. The experiment took approximately 45 minutes.

Data analysis—Data from the right eye were analyzed, using EyeDoctor and EyeDry analysis software (Clifton, Straccuzi, & Kinsey, 2006). An automatic routine combined fixations that were less than 80 ms with a previous or subsequent fixation that was within one character. Following that, fixations less than 80 ms or greater than 1000 ms were excluded.

We report analyses for all regions of the sentences, denoted by square bracketing: [The caterer tested][the filling][that was in][the tooth][before][leaving.] We label the regions as *Context NP + Verb*, *Polysemous NP*, *Between*, *Disambiguating*, *Spillover*, and *Wrap-up*. We combined the Context NP with the Verb because the probability of a first pass fixation was too low for the Context NP alone in the neutral context conditions (a pronoun, *M* = 33%), but was acceptable when combined with the Verb (100%). The Polysemous NP region was the determiner and noun, the Between region was the next 2–4 words, the Disambiguating region was the target noun or two-word phrase, and the Spillover region was defined as the next 5–7 characters, which was either one word or two short words. The Wrap-up region was the remaining material. The online supplement presents the regions of each item.

Two main eye-tracking measures are reported. *Regression Path* duration (also called *Go-Past Time*) is the sum of time from when a reader first fixates within a target region until the reader fixates anything to the right of the region, which includes first pass time and re-inspection of prior regions. This is a relatively early processing measure. *Second Pass* time is the sum of all rereading in a region after having previously exited the region to the right. This is considered a late processing measure.

Analyses for three additional measures are also reported when applicable. *First Pass* time is the sum of all fixations inside a region beginning with the first fixation inside until the gaze travels outside the region, either to the left or right, given that the reader has not yet fixated subsequent text to the right. (For convenience, we use First Pass to refer to both one word and multi-word regions.) *Regressions Out* is the probability of leaving a region on the saccade immediately following a first pass fixation to regress to earlier parts of the sentence. *Total Time* is the sum of all fixations in a region. Descriptive and inferential results for these three measures are reported in the online supplement.

Participants' overall accuracy on the comprehension questions was 94%. We omitted trials with major tracker loss or excessive blinking (5.2% of trials). On the basis of the First Pass measure, trials with skips in consecutive regions were excluded for all measures (1.6% of remaining trials). In total, 6.8% of the data were excluded. Any remaining skips of a region for a given measure were treated as missing data points, except Second Pass and Regression Path durations, which included zero times for that particular measure (Pickering et al., 2004). Data analyses were similar to those performed in Experiment 2.

Results and Discussion

Table 6 shows the two main eye-tracking measures for the six regions, and Table 7 summarizes the inferential tests. As in the previous experiments, the shorter neutral context was read more quickly than the longer biasing contexts, and is not considered further. On the polysemous NP region there were no early measure differences. Second Pass and Total Time showed a main effect of sense, with shorter durations for the dominant sense compared to the subordinate sense. A 3 (context) × 2 (sense completion) interaction was found for several later regions, appearing mainly when the 2 × 2 interaction for biasing context was also present, discussed in the next section.

Effects of biasing context—Overall, the 2 × 2 interactions indicate that inconsistent conditions were more difficult to process than consistent ones. Furthermore, in the inconsistent conditions, dominant context followed by the subordinate sense was more difficult to resolve than subordinate context followed by the dominant sense.

Readers regressed from the Spillover and Wrap-up producing an interaction between the four conditions for the Regression Path measure. In particular, following dominant context it was significantly more difficult to process the inconsistent subordinate sense than the consistent dominant sense. Also, for the Regressions Out measure in the Wrap-up region, following subordinate context readers regressed out significantly more often after encountering the inconsistent sense than the consistent sense. There were no reliable First Pass differences for the biasing context conditions.

Second Pass reading times showed a much more widespread indication of difficulty in the inconsistent conditions, as the 2×2 interaction occurred for all regions. Figure 2 illustrates this pattern, showing that the consistent conditions, with the solid black lines, were faster than the inconsistent conditions, with the dashed black lines. Following dominant context, readers had greater difficulty with the subordinate sense than the dominant one. On the other hand, it was not quite as difficult to switch from a subordinate context to the dominant sense, as those differences did not reach full significance. The Total Time dependent measure produced the same pattern of findings as Second Pass.

Neutral context—There was little evidence of an early difference between the two neutral context conditions. Regression Path times and Regressions Out did not differ significantly in any region. In the First Pass measure, readers tended to read through the sentence until the end, leading to a marginal difference in the Wrap-up region where the subordinate sense condition had longer reading times than the dominant sense.

For Second Pass, we found that the subordinate sense was more difficult in the Disambiguating region (marginal by subjects). In addition, readers spent marginally more total time reading the Polysemous NP in the subordinate sense condition. Since there were no differences in the early measures for these regions, this result suggests that participants reread the sentence's beginning more after encountering a subordinate sense continuation.

We also tested whether the neutral context-dominant sense condition was processed any differently than the dominant context-dominant sense condition. We found only marginal indications that the dominant context condition was read more quickly than the neutral context condition during a First Pass, in two regions: the Between region, t_1 (53) = 1.78, p = .08, t_2 (24) = 1.51, p = .14, and the Disambiguating sense region, t_1 (53) = 1.71, p = .09, t_2 (24) = .99, p = .33. However, we also found that the dominant context condition produced significantly *more* Regressions Out of the Disambiguating sense region, t_1 (53) = 2.31, p =. 02, t_2 (24) = 2.05, p = .05, indicating that the First Pass and Regressions Out measures were probably trading off, in sum producing no difference between these two conditions. Figure 2 illustrates the close similarity of these two conditions for Second Pass reading times, as well (filled squares and filled circles). We conclude that the neutral context-dominant sense condition did not show any reliable difficulty in processing compared to the dominant context-dominant sense condition. This processing similarity is consistent with the separate sense account, which predicts that the dominant sense is generally committed to as a default sense.

Sense Frequency and Similarity—As the polysemous words exhibited large variability in sense frequency, we again used linear mixed-effects regression to ascertain whether sense frequency contributed to processing times in the two Neutral Context conditions. This method also allowed us to further demonstrate that sense similarity is a different factor than sense frequency. As before, we included Sense Completion, Dominance Score, and Similarity Rating, in all models for all regions. We also included Region Length for the Disambiguating region, since the conditions were not always the same length. Table 8 presents the results for the two main measures. Results for the three other measures are in the online supplement; significant effects are discussed in the text. The Between region did not exhibit any significant effects.

On the polysemous noun, we found an unexpected interaction of Dominance Score and Similarity, which was significant for all three early measures (First Pass, Regression Path, and Regressions Out) as well as Total Time. Polysemes with a highly dominant sense and high sense similarity posed greater difficulty for readers. We interpret this to indicate that there is greater immediate competition between the two senses when both sense dominance and similarity are strong. In a later measure, the polyseme also produced a main effect of similarity, with high similarity leading to shorter Second Pass reading times. Total Time also showed a main effect of sense completion, with longer times for the subordinate conditions.

On the Disambiguating region, the interaction of Sense Completion and Dominance Score had a significant effect for the early measures of Regression Path and Regressions Out. As predicted, and consistent with Experiment 2, with a greater frequency spread between the dominant and subordinate senses, comprehenders spent more time reading the subordinate sense. The Regression Path measure also showed a significant interaction between Sense Completion and Similarity, indicating that greater similarity of senses led to the subordinate continuation being read more than the dominant one. There was also a main effect of Region Length for several measures. The two early measures of Regression Path and First Pass seemed to trade off, where a longer region produced longer first pass times but shorter regression path times; a similar tradeoff occurred for the second pass and total time measures.

On the Spillover region, the interaction between Sense Completion and Dominance Score persisted, occurring for the early measures of Regression Path and Regressions Out, as well as the later Second Pass measure. The Sense Completion by Similarity Rating interaction also continued, appearing in Regression Path and Regressions Out. Finally, there was also a marginal main effect of Similarity for First Pass, with greater similarity producing faster reading times.

Summary—For the biasing context conditions, readers had more difficulty with the inconsistent conditions than the consistent ones. Context that instantiated the dominant sense of a polysemous noun was more difficult to overcome and reanalyze than the subordinate context. This frequency effect appeared relatively early at the Spillover region in the Regression Path measure and persisted in Second Pass and Total Time measures over widespread regions.

For the neutral context conditions, readers had some early difficulty with subordinate sense continuations, with marginally longer First Pass reading of the sentences in the Wrap-up region. In later measures, this dominance effect was relatively localized, appearing in the Disambiguating region for Second Pass reading time. Further analyses indicated that without biasing context preceding a polysemous noun, readers chose the dominant sense and committed to it just as readily as when the preceding context biased readers to expect the dominant sense of the noun, providing further evidence against a core representation.

The regression analyses also provided evidence of dominance effects in the neutral context conditions. In the Disambiguating and Spillover regions, the sense completion by dominance score interaction was significant for both early and later measures. Furthermore, sense similarity affected measures separately from the Dominance Score factor, by in large. The one caveat to the factors' independence was localized on the polyseme in early and late measures, and our interpretation of sense competition bears further investigation. In sum, Experiment 3 showed that sense frequency does affect processing of both the biased and neutral context conditions, and provides a finer-grained picture indicating that sense dominance does have early and later effects on the processing of polysemes, and is graded by degree of sense dominance in the neutral conditions.

General Discussion

We have focused on the role of sense frequency in these experiments, which was the tool used in prior sentence-processing research to draw conclusions about the representation of polysemous words (Frazier & Rayner, 1990; Frisson & Pickering, 1999; Pickering & Frisson, 2001). Using the same polysemous words as Klein and Murphy (2001), we examined how stimulus materials and reading tasks may have played a role in producing findings taken to support a common, underspecified representation in that previous work. We first summarize our most significant findings, supporting the view that polysemous words have separate sense representations. We then turn to processing issues that our results shed some light upon.

First, in all three experiments, continuations consistent with the contextually supported sense of the polysemous word were far easier to interpret than inconsistent ones. This indicates that our materials were successful in instantiating the dominant and subordinate senses intended and also that the correct sense of polysemous words can be easily derived with strong prior context (as in Frisson & Pickering, 1999). Second, when readers encountered an inconsistent continuation following a polysemous word, they found it more difficult to recover from preceding dominant-bias context than subordinate-bias context. This pattern, found in all experiments, shows that comprehension is sensitive to sense frequencies.

Most importantly, following a neutral context, we found that the dominant sense was easier to access than the subordinate one. In Experiment 1, when the polysemous noun and target sense appeared in different sentences, in different self-paced displays, this pattern was strong and reliable. In Experiment 2, when the polysemous noun and target sense appeared in the same sentence but in different self-paced moving window displays, this pattern was weaker but still apparent. In Experiment 3, the dominance effect was marginal in both early and later eye-tracking measures.

The varying strength of the dominance effect in the neutral context conditions bears further examination. While we found evidence supporting the dominant sense preference, the reading time differences in Experiment 2 were not as large as in Experiment 1. Frisson (2009) briefly reports a similar finding, suggesting that sentence boundaries have an effect on sense resolution. We consider three possibilities for the varying strength of the dominance effect. First, the sentence boundary could have encouraged readers to select a sense before continuing to read. In Experiments 2 and 3, readers may not have finished selecting a particular interpretation of the polysemous word by the disambiguating region. Bolstering this possibility are the stronger effects for the joint region than disambiguating or spillover regions alone in Experiment 2, and finding a First Pass difference in Experiment 3 only at the final region of the sentence. For example, consider the sentence *She counted the sheets left in the printer tray in the afternoon*. It is possible that the dominant (bed sheets) sense of *sheets* was not completed by the time disambiguating information was encountered, pushing the effect into the final phrase. In Experiment 1, however, the sentence boundary that followed the polysemous word might well have caused readers to complete the selection process before continuing to read.

Second, one factor is certainly that in Experiment 1 the reading time was taken across the whole second sentence, which might reflect more global discourse integration effects across the two sentences, whereas in Experiment 2, the target areas constituted smaller amounts of text. The dominance effect found in Experiment 3 for the re-reading measures supports this point. Third, a more interesting consequence of the different presentation procedures is that using self-paced reading with a moving window may have decreased the strength of the dominance effect in Experiment 2. In Experiment 1, reading times for the disambiguating sentence could have been longer partially because readers still had visual access to the disambiguating sense phrase (*The fabric…*), and could reinspect it during reanalysis. In contrast, Experiment 2's phrase-by-phrase presentation method did not allow re-reading of previous material, instead requiring readers to rely on memory representations if information was to be reaccessed or reanalyzed. This may have prevented or degraded reanalysis of incorrect senses in some cases and therefore reduced average reading times. That the dominance effect in Experiment 3 appeared at the Disambiguating region in the Second Pass measure supports this interpretation.

Finally, we note that in all three experiments, analyses for the neutral conditions with the stimulus items as a random factor showed weaker dominance effects than those with participants as a random factor; in Experiment 1 the dominance effects were small, but generally reliable, but in Experiments 2 and 3, the difference between neutral conditions did not always reach significance over item means. Further analyses, however, showed that this was likely caused by the inclusion of polysemes with near equi-biased senses, as the size of the dominance score predicted the reading time advantage for the neutral context-dominant sense vs. neutral context-subordinate sense conditions in Experiments 2 and 3 (the Sense Completion by Dominance Score interaction). All these results, then, support the conclusion that readers tend to interpret a polysemous word in its more frequent sense, but primarily when the frequency differences are large.

Comparison to Prior Results

Our results seem to conflict with many past statements in the literature concerning underspecification of polysemous words, which argue that when there is no biasing context, readers retrieve only a minimal word sense rather than a full-fledged meaning. However, an examination of the specific findings of the literature shows that there are in fact fewer empirical conflicts between our results and past experiments than the differing conclusions would suggest. We briefly consider how our results compare to relevant past studies, focusing only on the aspects of these studies that relate to our concerns—the experiments also had other goals and interesting effects we don't consider.

Frisson and Pickering (1999) provided a number of detailed arguments for underspecified senses of these words, but their study did not have a neutral condition that tested whether one sense would be activated more than another. As well, we found a sense frequency effect for the inconsistent biasing context conditions, which they did not find in a post hoc analysis.

Pickering and Frisson (2001) tested homonymous and polysemous verbs, in separate experiments. They compared preferred to nonpreferred interpretations (sense dominance) and did have neutral conditions, although they did not test inconsistent conditions. Their results for polysemous verbs are very similar to our Experiment 3 results, in terms of the measures, the regions, and the conditions showing dominance effects. Furthermore, they found dominance effects for both kinds of verbs, but only in later reading measures (Second Pass, etc.). The results for polysemous verbs in that study are very similar to our present results with polysemous nouns.

Although those prior studies disagreed with our conclusions about sense representation in various respects, often based on theoretical arguments we do not consider here, the only strong *empirical* conflict with the present results is Frazier and Rayner's (1990) failure to find a dominance effect for polysemous words in neutral contexts. Therefore, we next discuss the differences between their and our study that might explain this disparity.

One important aspect of our design was the use of unambiguous conditions. Our experiments used three contexts: a neutral context and contexts biased towards the dominant or subordinate sense of a polysemous word. We demonstrated that when the prior context clearly picked out a given sense, comprehension of the subsequent sentence or phrase was strongly affected, that is, consistent context-to-sense conditions were easier than inconsistent ones. Frazier and Rayner (1990) used consistent and neutral contexts, but not inconsistent contexts. With such a design, one does not know whether initially choosing the unintended sense *would* produce a strong effect on reading the rest of the sentence. If the disambiguating region of our materials was only weakly *in*compatible with the unintended sense, a null effect could result, even if readers had earlier chosen the other, intended sense.

For example, Frazier and Rayner's sentence *The records were carefully guarded after the political takeover* plausibly implies the reports or files sense of *records*, but it does not contradict the sound recording sense. There is nothing incompatible with sound recordings being guarded, so it is possible that readers could integrate the disambiguating *political takeover* region with whichever sense of the word had been activated. We know that this issue is a concern, because in an initial experiment not reported here, we found weak effects in the biasing context conditions (such that context biasing the dominant sense of the word did not greatly slow the reading of a subordinate sense continuation), accompanied by an even weaker dominance effect after neutral context. When we revised and pretested the continuations to create a strong effect in the biasing context conditions, we then also obtained the dominance effect following neutral context. In contrast, Frazier and Rayner reported null effects of dominance for polysemous words without simultaneously reporting positive results of contextually constrained uses of those words with their stimuli. To interpret such null results conclusively, it would be useful to demonstrate positive effects in biasing contexts, as the present study has.

Finally, the differences between previous studies finding no effect of sense dominance and our own finding of this effect could well be due to differences in the items used. A major motivation for the present experiments was to test the stimuli that have suggested that polysemous senses are distinct (Klein & Murphy, 2001, 2002), using a sentence-reading task. Given that the sensicality judgments, conceptual judgments, and now dominance effects in reading all indicate that these words' senses are distinct, this suggests that differences between the present set of stimuli and those used in previous reading experiments could be important. We address this issue below.

Sense Representation and Processing

The present study addressed the question of whether readers access an underspecified, core meaning representation when they encounter a polysemous word and then wait for further information before selecting (or constructing) a specific sense. Our finding a dominance effect in neutral contexts is evidence against such an account and in particular rules out the strong claim that polysemous senses are not represented in the lexicon at all but are derived in context (e.g., Caramazza & Grober, 1976; Nunberg, 1979). If different senses were not already represented in the lexicon, then there could not be any advantage to reading a continuation consistent with the dominant sense, since it would not have been activated after reading the word in a neutral context. Nor should reanalysis following an inconsistent dominant context be harder than following a subordinate context.

A somewhat weaker claim, the partial specification hypothesis, was made by Frazier and Rayner (1990) and Pickering and Frisson (2001), who suggested that readers may hold off on selecting a sense until there is clear evidence for which one is correct. They suggest that this is possible for polysemous words because of semantic features that are common to the senses (see also Frazier, 1999). Although we disagree with the latter assumption about representation, our results are generally consistent with their proposal of delayed or slower processing. Although we found that readers made a commitment to a single sense of a polysemous word, this was clearly stronger at the end of the sentences than in the middle. This finding is in contrast with the processing of homonyms, where readers very quickly choose an interpretation, even when there is no biasing context (e.g., Duffy et al., 1988; Frazier & Rayner, 1990). However, our results are not as consistent with the idea of an underspecified representation, because in all three experiments it was just as easy for readers to access the dominant sense of a word without supportive context (neutral) as it was to access it with supportive context (dominant). If readers had selected an underspecified sense, it is surprising that they did not then slow down when filling the meaning out when it was disambiguated (Frazier & Rayner, p. 190). We cannot make too much of this (consistently obtained) null effect, but when it is combined with the positive effect of sense dominance, it does suggest that readers are selecting the dominant sense to a measurable degree even when there is no biasing context.

In summary, the results from all three experiments are consistent with the idea that readers represent the different senses of polysemous words rather than a core meaning, because the more frequent sense is treated as a default meaning, and as the strength of the dominant sense increased, the greater the difficulty in reading the subordinate continuation. However, there is still much to be learned about how those senses are accessed and the time course of that process.

Different Forms of Polysemy

The question of why we found dominance effects for polysemous words whereas other studies have not raises the question of possible item differences, which is in turn related to a theoretically important issue—whether different forms of polysemy are represented and processed in the same way. Some examples of polysemy come about through productive patterns or lexical rules. For example, one can generally use the same word to refer to an animal and its meat: *I saw a lamb/salmon/penguin at the zoo* versus *Would you like to try some roast lamb/salmon/penguin?* One can often use the same word to refer to a container and to the amount that container holds: *I cracked the cup/bucket/bowl* versus *I ate a cup/bucket/bowl of soup*. However, other forms of polysemy are idiosyncratic and can be found in only one or a very few lexical items, for example, *The city's atmosphere was polluted* versus *The restaurant's atmosphere was relaxed*.

Linguists have argued that productive forms of polysemy may be quite different from idiosyncratic examples, though usually because one type is susceptible to linguistic analysis and the other is not, rather than because of psycholinguistic principles or empirical results (e.g., Copestake & Briscoe, 1995; Pustejovsky, 1995; cf. Murphy, 2007). One possibility is that the relation between senses in productive forms may allow readers to negotiate the different senses. That is, readers may not commit to one particular sense of a word, but instead activate two related senses of a polysemous word, in parallel, and the principled relation between those two senses allows a relatively easy selection between them without measurable processing cost. Indeed, Frisson and Pickering (1999) note that their results are also consistent with a parallel model that activates both the literal and metonymic senses of a noun. For example, *library* has a literal or concrete "building" sense, as in *The father strolled to the library*, as well as an abstract or metonymic "institution" sense, as in *The father complained to the library*. Note that one sense of the word implies that the other sense of the word exists. The bi-directional relationship between these two senses may therefore provide a low-cost way to transition between two such senses, both of which may appear in the same discourse.

These different possibilities may furthermore depend on the nature of the polysemic senses. Klepousniotou et al. (2008) argued that priming vs. competition of polysemic senses depends on how semantically similar the senses are, suggesting that the Klein and Murphy stimuli were on the whole fairly dissimilar. If this is correct, then it is possible that some polysemes have separate representations and others do not. Testing this possibility is tricky, however, because as the senses become more and more similar, it is increasingly difficult to know that they are in fact different senses, and it is also more difficult to be sure that the experimental contexts adequately distinguish them.[1] Nonetheless, the semantic overlap of the senses may have an effect on switching between senses, as Klepousniotou et al.'s results suggest[2]. Our own results found faster reading for polysemes with more similar senses, in addition to sense dominance effects. Because our study was not designed to test this variable, we only conclude that our finding of a dominance effect was not caused by sense similarity.

However, we are not so sure that even senses judged as similar have a core meaning that is consistent across contexts. In many such proposals in the psycholinguistic literature, no example core meaning is ever specified. Linguists who have attempted to discover core meanings of highly polysemous words have generally failed to do so (e.g., Rice, 1992). Klepousniotou et al. (2008, p. 1535), then, are to be complimented for actually specifying, as an example, what they believe to be the core meaning of the word *rabbit* used in its animal and meat senses. They propose that the features "animate," "farm animal," "edible," and "meat" apply to most uses of rabbit and therefore could be the initial representation when the word is encountered. However, we would argue that these features are in fact not present in both senses of *rabbit*. When referring to a hopping animal, the features of edibility and meat are generally absent. A living rabbit is not edible except in the hypothetical sense that one kills it, skins it, and prepares it first—i.e., turns it into meat, which is the other sense of the word. Similarly, *I'm cooking rabbit* does not refer to an animate farm animal. Also, many polysemous words have multiple senses—not merely two—making it more difficult to specify a core meaning that is consistent with all of them. For example, *rabbit* has senses such as rabbit fur (*She wore rabbit*), a fast runner, and a mechanical device chased by greyhounds. None of these is edible, a farm animal, or meat.

Indeed, there is little semantic overlap in many forms of polysemy tested in the literature, which often involve ontologically very different referents, such as place/institution (like *library*), place/event (like *Vietnam*), object/substance (*rabbit, oak, cotton*), representational object/content (*novel, film*), and so on. The properties of events are very different from those of locations; and the propositional content of a novel has none of the physical properties of the physical novel, even though participants rate the two as highly related.

For these reasons, if it is the case that some polysemous words yield no (or delayed) dominance effects, then we suspect that this is not due to activation of an underspecified or core meaning that is compatible with (all) the senses, but that instead, there may simply be little interpretation at all until the later context arrives or the sentence ends. It is also possible that all the senses are activated in parallel in some of these cases, which would explain the lack of a frequency effect in some experiments. We agree with Frisson (2009) that the large

[1]To pick one example, Klepousniotou et al.'s first-listed polysemous word with high semantic overlap was *article*. The two senses were distinguished by modifiers, as in *well-written article* vs. *popular article*. Presumably, the first refers to the content and the second to instantiation of the article, but it is also possible that it is the content that is popular. Thus, there is the danger that "similar" senses can turn into identical senses in the materials. It was to avoid such methodological issues that Klein & Murphy (2001) chose words with fairly distinct senses.

[2]Klepousniotou et al. consistently refer to their comparison as involving semantic "overlap," but their instructions apparently asked people to rate the "relatedness" of senses, which may tap into a number of relations besides feature overlap. Our analyses in Experiments 2 and 3 therefore used similarity judgments.

number of senses for many frequent words makes full parallel activation unlikely, but our regression results suggest that graded parallel activation of competing senses, activated by frequency and contextual information, is a processing strategy that could yield delayed dominance effects (a reordered access model, Dopkins, Morris, & Rayner, 1992; Duffy et al., 1988; Sheridan, Reingold, & Daneman, 2009). This possibility is a fruitful direction for future work.

Most past research has either tested a broad range of polysemous words (as in our work) or has focused on very specific forms of polysemy. Each approach has its advantages and disadvantages, but the conflicts in the literature suggest that it would be useful to broaden future studies by systematically comparing distinct classes of polysemy. Rather than admitting any form of polysemy into an undifferentiated set of items and rather than testing only one specific type, it would be informative to compare within a single experiment different types that might plausibly differ in their representation or processing (as Klepousniotou et al., 2008, did). We did this to some degree in our regression analyses of sense similarity and dominance. Future studies could examine whether the sense frequency effects we found might be more important for one type of polysemy than another—say, for idiosyncratic forms more than for productive, rule-derived forms. Similarly, a more systematic investigation of the relations between different senses—such as their similarity or specific thematic relations joining them—may provide further important information on how polysemous words are processed. It is possible that questions about how senses are activated do not have a single answer but differ depending on the word and the nature of the polysemy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by NIMH grant MH41704. We thank Kim Larrimore, Robyn Kim, Andrea Martin, and Julie Van Dyke for their help with data collection, and Doug Roland and Hong Oak Yun for assistance with the LMER analyses. We thank Brian McElree, Steven Frisson, Keith Rayner, and two anonymous reviewers for helpful comments on previous versions of this manuscript.

References

Baayen, RH. R package version 1.4. 2011. languageR: Data sets and functions with "Analyzing linguistic data: A practical introduction to statistics".

Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language. 2008; 59:390–412.

Bates, D.; Maechler, M.; Bolker, B. lme4: Linear mixed-effects models using S4 classes. 2011. Version 0.999375–42

Caramazza, A.; Grober, E. Semantics: Theory and application. Washington, DC: Georgetown University Press; 1976. Polysemy and the structure of the subjective lexicon. In C. Rameh (Ed.), Georgetown University roundtable on languages and linguistics; p. 181-206.

Clark HH. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior. 1973; 12:355–359.

Clifton, C.; Straccuzi, D.; Kinsey, J. EyeDoctor (Version 0.6.3g) and EyeDry (Version 8/6/02) [Computer Software]. 2006. Available from http://www.psych.umass.edu/eyelab/software/

Copestake A, Briscoe T. Semi-productive polysemy and sense production. Journal of Semantics. 1995; 12:15–67.

Cruse, DA. Lexical semantics. Cambridge: Cambridge University Press; 1986.

Dopkins S, Morris RK, Rayner K. Lexical ambiguity and eye fixations in reading: A test of competing models of lexical ambiguity resolution. Journal of Memory and Language. 1992; 31:461–476.

Duffy SA, Morris RK, Rayner K. Lexical ambiguity and fixation times in reading. Journal of Memory and Language. 1988; 27:429–446.

Francis, WN.; Kucera, H. Frequency analysis of English usage: Lexicon and grammar. Boston: Houghton Mifflin; 1982.

Frazier, L. On sentence interpretation. Dordrecht: Kluwer Academic Publishers; 1999.

Frazier L, Rayner K. Taking on semantic commitments: Processing multiple meanings vs. multiple senses. Journal of Memory and Language. 1990; 29:181–200.

Frisson S. Semantic underspecification in language processing. Language and Linguistics Compass. 2009; 3:111–127.

Frisson S, Frazier L. Carving up word meaning: Portioning and grinding. Journal of Memory and Language. 2005; 53:277–291.

Frisson S, Pickering MJ. The processing of metonymy: Evidence from eye movements. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1999; 25:1366–1383.

Gelman, A.; Hill, J. Data analysis using regression and multilevel/hierarchical models. New York, NY: Cambridge University Press; 2007.

Jaeger, TF. Random effect: Should I stay or should I go? [Web blog post]. 2009 May 14. http://hlplab.wordpress.com/2009/05/14/random-effect-structure/Retrieved August 1, 2011

Klein DE, Murphy GL. The representation of polysemous words. Journal of Memory and Language. 2001; 45:259–282.

Klein DE, Murphy GL. Paper has been my ruin: Conceptual relations of polysemous senses. Journal of Memory and Language. 2002; 47:548–570.

Klepousniotou E, Titone D, Romero C. Making sense of word senses: The comprehension of polysemy depends on sense overlap. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2008; 34:1534–1543.

Lakoff, G. Women, fire and dangerous things. Chicago: University of Chicago Press; 1987.

Murphy, GL. Parsimony and the psychological representation of polysemous words. In: Rakova, M.; Petho, G.; Rákosi, C., editors. The cognitive basis of polysemy. Frankfurt am Main, Germany: Peter Lang Verlag; 2007.

Nunberg G. The non-uniqueness of semantic solutions: Polysemy. Linguistics and Philosophy. 1979; 3:143–184.

Pickering MJ, Frisson S. Processing ambiguous verbs: Evidence from eye movements. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2001; 27:556–573.

Pickering, MJ.; Frisson, S.; McElree, B.; Traxler, MJ. Eye movements and semantic composition. In: Carreiras, M.; Clifton, JC., editors. The on-line study of sentence comprehension: Eyetracking, ERPs, and beyond. Hove: Psychology Press; 2004. p. 33-50.

Pustejovsky, J. The generative lexicon. Cambridge, MA: MIT Press; 1995.

Pylkkänen L, Llinas R, Murphy GL. The representation of polysemy: MEG evidence. Journal of Cognitive Neuroscience. 2006; 18:97–109. [PubMed: 16417686]

Rayner K. Eye movements in reading and information processing: 20 years of research. Psychological Bulletin. 1998; 124:372–422. [PubMed: 9849112]

Rayner K, Duffy S. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Memory & Cognition. 1986; 14:191–201.

Rice, SA. Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum; 1992. Polysemy and lexical representation: The case of three English prepositions; p. 89-94.

Rodd JM, Gaskell MG, Marslen-Wilson WD. Modelling the effects of semantic ambiguity in word recognition. Cognitive Science. 28:89–104.

Ruhl, C. On monosemy: A study in linguistic semantics. Albany: State University of New York Press; 1989.

Schneider, W.; Eschman, A.; Zuccolotto, A. E-Prime (Version 1.1). Pittsburgh, PA: Psychology Software Tools; 2002.

Sheridan H, Reingold EM, Daneman M. Using puns to study contextual influences on lexical ambiguity resolution: Evidence from eye movements. Psychonomic Bulletin & Review. 2009; 16:875–881. [PubMed: 19815792]

Stracuzzi, D.; Kinsey, J. EyeTrack (Version 0.6.x) [Computer software]. 2006. Available from http://www.psych.umass.edu/eyelab/software/

Swinney DA. Lexical access during sentence comprehension: (Re)consideration of context effects. Journal of Verbal Learning and Verbal Behavior. 1979; 18:645–659.

Tanenhaus MK, Leiman JM, Seidenberg MS. Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. Journal of Verbal Learning and Verbal Behavior. 1979; 18:427–440.

Zgusta, L. Manual of lexicography. The Hague: Mouton; 1971.

Highlights

- Three experiments showed sense frequency effects when processing polysemous words.
- more difficulty shifting from dominant context to subordinate sense than the reverse after neutral context, dominant sense read more quickly than subordinate
- The stronger a dominant sense is, the greater the sense frequency effect found.
- Results support separate senses versus an underspecified or core representation.

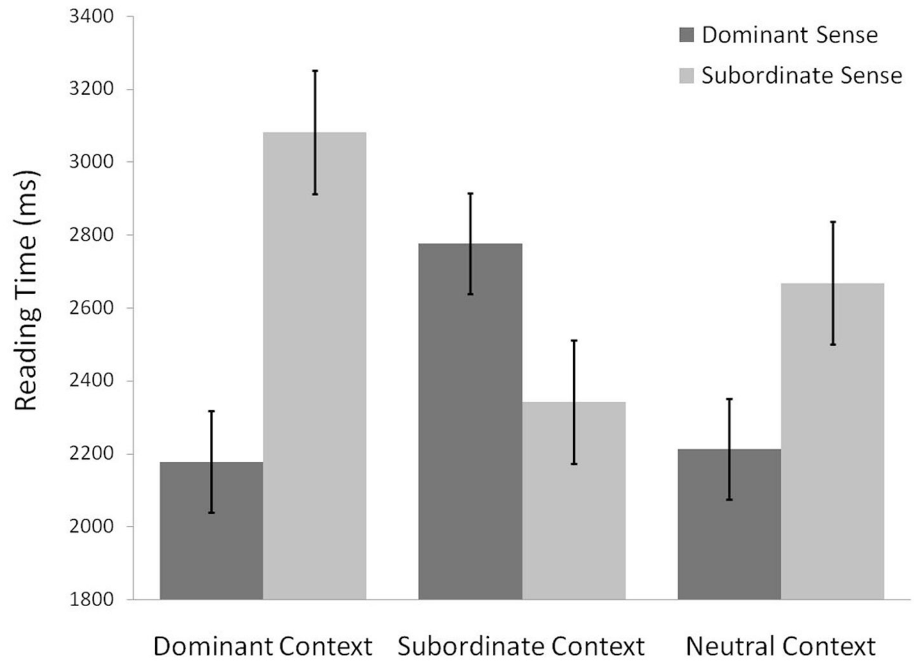
3400
3200
3000
2800
2600
2400
2200
2000
1800
Reading Time (ms)
Dominant Sense
Subordinate Sense
Dominant Context
Subordinate Context
Neutral Context

Figure 1.

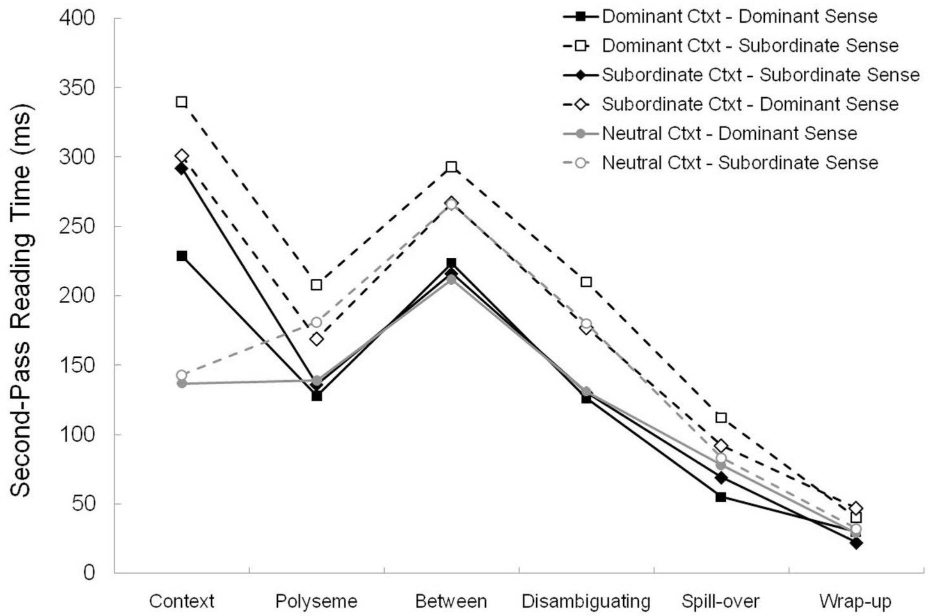
Dominant Ctxt - Dominant Sense
Dominant Ctxt - Subordinate Sense
Subordinate Ctxt - Subordinate Sense
Subordinate Ctxt - Dominant Sense
Neutral Ctxt - Dominant Sense
Neutral Ctxt - Subordinate Sense
Second-Pass Reading Time (ms)
400
350
300
250
200
150
100
50
0
Context
Polyseme
Between
Disambiguating
Spill-over
Wrap-up

Figure 2.

Table 1

Example Materials for Experiments 1, 2, and 3.

Context	Sense Completion	Sentence(s)
		Experiment 1
Dominant	Dominant	The fashion designers discussed the cotton. The fabric was not what they had been hoping for.
Dominant	Subordinate	The fashion designers discussed the cotton. The crop was not what they had been hoping for.
Subordinate	Dominant	The farm owners discussed the cotton. The fabric was not what they had been hoping for.
Subordinate	Subordinate	The farm owners discussed the cotton. The crop was not what they had been hoping for.
Neutral	Dominant	They discussed the cotton. The fabric was not what they had been hoping for.
Neutral	Subordinate	They discussed the cotton. The crop was not what they had been hoping for.
		Experiments 2 & 3
Dominant	Dominant	The fashion designers discussed the cotton after the fabric ripped a second time.
Dominant	Subordinate	The fashion designers discussed the cotton after the crop failed a second time.
Subordinate	Dominant	The farm owners discussed the cotton after the fabric ripped a second time.
Subordinate	Subordinate	The farm owners discussed the cotton after the crop failed a second time.
Neutral	Dominant	They discussed the cotton after the fabric ripped a second time.
Neutral	Subordinate	They discussed the cotton after the crop failed a second time.

Table 2

Sense Dominance and Sense Similarity Ratings for the Polysemous Words in Neutral Contexts

Word	Percent dominant	Dominant sense gloss	Percent subordinate	Subordinate sense gloss	Sense Similarity
Coat	91	outer garment	1	layer or covering	2.81
Orange	89	citrus fruit	1	a bright reddish-yellow hue	2.94
Book	89	informational content of a printed volume	3	physical object consisting of printed pages and a binding	6.56
Sign	86	publicly displayed board with information	0	symbol used in algebra, music, etc.	3.00
Drinkers	83	consumers of alcoholic beverages	0	consumers of liquid	5.19
Oak	83	woody, tall plant	4	wood used for building	6.20
Drive	81	journey in an automobile	0	charitable effort	2.15
Boxes	81	containers with flat sides	1	separate compartments for groups of people	3.69
Corn	76	type of food	14	type of crop	6.19
Atmosphere	74	envelope of gases surrounding a planet	10	feeling of a situation	3.56
Trunk	70	a case with a hinged lid	0	main stem of a tree	1.38
Production	68	a film, play, or record	12	process of being manufactured	3.81
Tin	66	a small, lidded receptacle	8	silvery, malleable substance	3.38
Cold	66	low degree of heat in the air	16	illness causing runny nose and cough	2.13
Letters	64	written or printed communication	8	character in the alphabet	2.63
Paper	58	thin sheets used for writing, printing, or wrapping	17	newspaper	4.38
School	42	an educational institution	40	a building that houses an educational institution	6.00
Filling	52	center of pastry or other food	29	material that occupies a cavity in a tooth	3.94
Cotton	51	type of thread or cloth	31	plant containing a soft white substance	5.63
Glasses	52	spectacles	14	drinking vessels	2.94
Chicken	53	type of food	40	domestic fowl	6.31
Shower	51	bathing under a spray of water	36	the apparatus used to bathe	5.88
Fortune	48	chance	33	riches	2.88
Classes	42	school courses	7	divisions of society	2.00
Sheets	27	pieces of paper for writing	19	large pieces of cloth used for bedding	1.88

Note. Sense similarity ratings were on a scale from 1(completely different) and 7 (almost identical).

Table 3

Reading Times (ms) for Experiment 2

Condition		Region				
Context	**Target Sense**	**Context**	**Verb**	**Polysemous NP**	**Disambiguating**	**Spillover**
Dominant	Dominant	833 (49)	785 (51)	799 (59)	981 (61)	929 (63)
Dominant	Subordinate	816 (53)	764 (60)	788 (55)	1173 (89)	1160 (83)
Subordinate	Dominant	775 (56)	739 (45)	795 (62)	1045 (63)	1103 (77)
Subordinate	Subordinate	802 (66)	755 (43)	776 (50)	974 (64)	971 (63)
Neutral	Dominant	522 (24)	640 (36)	746 (56)	956 (61)	888 (59)
Neutral	Subordinate	556 (36)	709 (46)	782 (66)	1019 (73)	1027 (82)

Note. Standard error of the mean is shown in parentheses.

Table 4

Inferential Statistical Tests for Experiment 2

Analysis	Statistic	*df*	Disambiguating	Spillover
Context main effect	*F*1	2, 60	3.44*	2.90†
	*F*2	2, 48	3.08*	1.45
Sense main effect	*F*1	1, 30	6.27*	4.24*
	*F*2	1, 24	2.46	1.53
Interaction	*F*1	2, 60	9.41***	7.27***
	*F*2	2, 48	3.12*	2.72†
2×2 Interaction, of Biasing Context	*F*1	1, 30	16.14***	14.15***
	*F*2	1, 24	5.26*	4.04†
Dominant Context	*t*1	35	3.58***	3.44**
	*t*2	24	2.31*	2.05*
Subordinate Context	*t*1	35	1.80†	2.07*
	*t*2	24	1.11	1.07
Neutral Context	*t*1	35	1.74†	1.83†
	*t*2	24	1.03	1.47

Note. Dominant and Subordinate Context *t*-tests are for the inconsistent condition minus the consistent condition; Neutral Context *t*-tests are for the subordinate sense condition minus the dominant sense condition.

*** $p \leq .001$,

** $p \leq .01$,

* $p \leq .05$,

† $p \leq .10$

Table 5

Summary of Fixed Effect Predictors for Neutral Context Reading Times in Experiment 2

Region	Term	Estimated Coefficient	Standard Error	*t*-value
Disambiguating Phrase	(Intercept)	985.63	63.87	15.43
	Region Length	32.54	5.27	6.17
	Sense Completion	36.41	46.68	.78
	Similarity Rating	−30.64	17.05	−1.80
	Sense Completion * Dominance Score	273.47	148.68	1.84
Spillover	(Intercept)	732.56	126.53	5.79
	Region Length	10.93	9.81	1.11
	Sense Completion	110.53	50.41	2.19
	Similarity Rating	−53.46	22.05	−2.43
	Sense Completion * Dominance Score	200.78	161.29	1.25

Note. Differences with $t \geq 2.0$ are significant at $p \leq .05$.

Table 6

Eye-tracking Measures for Experiment 3: Two Main Measures

Measure	Condition		Region					
	Context	Sense	Context	Polyseme	Between	Disambiguating	Spillover	Wrap-up
Regression Path (ms)	Dominant	Dominant		373 (20)	435 (33)	437 (34)	373 (48)	374 (70)
	Subordinate	Dominant		409 (26)	434 (36)	450 (35)	526 (56)	666 (99)
	Subordinate	Subordinate		419 (30)	449 (31)	402 (29)	420 (63)	492 (92)
	Dominant	Subordinate		401 (23)	434 (31)	427 (25)	506 (67)	618 (100)
	Neutral	Dominant		383 (25)	449 (29)	405 (27)	434 (45)	435 (65)
	Neutral	Subordinate		371 (24)	466 (34)	438 (35)	521 (55)	494 (62)
Second Pass (ms)	Dominant	Dominant	229 (37)	128 (19)	224 (29)	126 (16)	55 (10)	30 (9)
	Dominant	Subordinate	301 (41)	169 (22)	267 (31)	177 (22)	92 (15)	47 (12)
	Subordinate	Dominant	340 (42)	208 (28)	293 (37)	210 (25)	112 (20)	40 (10)
	Subordinate	Subordinate	292 (52)	136 (21)	216 (32)	130 (26)	69 (15)	22 (8)
	Neutral	Dominant	137 (26)	139 (22)	212 (30)	131 (21)	78 (13)	29 (8)
	Neutral	Subordinate	143 (25)	181 (27)	266 (46)	180 (26)	83 (17)	32 (8)

Note. Standard error of the mean appears in parentheses.

Table 7

Inferential Tests for Experiment 3: Two Main Measures

Measure	Analysis	Statistic	*df*	Polyseme	Between	Disambiguating	Spillover	Wrap-up
Regression Path	Context main effect	*F*1	2, 88	1.41	< 1	1.41	< 1	1.25
		*F*2	2, 48	1.53	1.02	< 1	< 1	< 1
	Sense main effect	*F*1	1, 44	< 1	< 1	< 1	< 1	2.45
		*F*2	1, 24	< 1	< 1	< 1	1.17	1.34
	Interaction	*F*1	2, 88	< 1	< 1	< 1	2.27	5.75**
		*F*2	2, 48	< 1	< 1	< 1	1.76	4.92**
	2×2 Interaction, Biasing Contexts	*F*1	1, 44	< 1	< 1	< 1	3.45†	8.93**
		*F*2	1, 24	< 1	< 1	< 1	3.36†	7.16**
	Dominant Context	*t*1	53	1.30	< 1	< 1	2.08*	3.20**
		*t*2	24	1.22	< 1	< 1	2.20*	2.46*
	Subordinate Context	*t*1	52	< 1	< 1	< 1	< 1	1.10
		*t*2	24	< 1	< 1	< 1	< 1	1.42
	Neutral Context	*t*1	52	< 1	< 1	< 1	1.32	< 1
		*t*2	24	< 1	< 1	< 1	1.02	< 1
Second Pass	Context main effect	*F*1	2, 88	< 1	< 1	< 1	< 1	< 1
		*F*2	2, 48	< 1	< 1	< 1	< 1	< 1
	Sense main effect	*F*1	1, 44	3.73†	1.41	3.36†	1.53	< 1
		*F*2	1, 24	5.67*	2.15	1.31	1.24	< 1
	3 × 2 Interaction	*F*1	2, 88	4.41**	4.32*	6.39**	4.29*	1.67
		*F*2	2, 48	6.00**	2.35	7.19**	6.44**	2.35
	2×2 Interaction, Biasing Contexts	*F*1	1, 44	7.70**	8.88**	11.88***	7.90**	3.27†
		*F*2	1, 24	11.43**	3.97†	8.55**	14.16***	5.08*
	Dominant Context	*t*1	53	2.76**	2.33*	3.52***	2.74**	< 1
		*t*2	24	3.44**	1.97†	2.31*	3.40**	< 1

Measure	Analysis	Statistic	*df*	Polyseme	Between	Disambiguating	Spillover	Wrap-up
	Subordinate Context	*t*1	52	1.35	1.30	1.68†	1.25	1.66†
		*t*2	24	1.56	1.18	1.51	2.01†	1.75†
	Neutral Context	*t*1	52	1.34	1.29	1.69†	< 1	< 1
		*t*2	24	1.68†	1.35	2.51*	< 1	< 1

Note. Dominant and Subordinate Context *t*-tests are for the inconsistent condition minus the consistent condition; Neutral Context *t*-tests are for the subordinate sense condition minus the dominant sense condition.

*** $p \leq .001$,

** $p \leq .01$,

* $p \leq .05$,

† $p \leq .10$

Table 8

Summary of Fixed Effect Predictors in Experiment 3, Neutral Context Conditions: Two Main Measures

Measure	Region	Term	Estimated Coefficient	Standard Error	*t*-value
Regression Path	Polysemous NP	(Intercept)	370.14	19.80	18.69
		Dominance Score	24.26	51.98	.47
		Similarity Rating	2.18	9.97	.22
		Dominance Score * Similarity Rating	77.11	36.43	2.12
	Disambiguating Phrase	(Intercept)	438.13	35.86	12.22
		Region Length	−24.70	6.47	−3.82
		Sense Completion	8.44	40.87	.21
		Similarity Rating	−4.69	15.83	−.30
		Sense Completion * Dominance Score	277.71	129.54	2.14
		Sense Completion * Similarity Rating	60.42	28.62	2.11
	Spillover	(Intercept)	429.37	37.69	11.39
		Sense Completion	8.83	29.14	.30
		Similarity Rating	−13.73	22.02	−.62
		Sense Completion * Dominance Score	250.110	107.00	2.34
		Sense Completion * Similarity Rating	63.16	25.25	2.50
Second Pass	Polysemous NP	(Intercept)	158.91	19.53	8.14
		Sense Completion	50.24	32.01	1.57
		Similarity Rating	−26.55	9.58	−2.74
		Sense Completion * Dominance Score	24.92	89.08	.28
	Disambiguating Phrase	(Intercept)	158.29	23.01	6.88
		Region Length	12.03	4.02	2.99
		Sense Completion	52.07	29.49	1.77
		Similarity Rating	−12.48	11.01	−1.13
		Sense Completion * Dominance Score	85.79	79.44	1.08
	Spillover	(Intercept)	68.87	9.87	6.98
		Sense Completion	8.84	16.18	.55

Measure	Region	Term	Estimated Coefficient	Standard Error	*t*-value
		Similarity Rating	4.68	6.41	.73
		Sense Completion * Dominance Score	101.07	48.52	2.08

Note. Differences with $t \geq 2.0$ are significant at $p \leq .05$.