Journal of Experimental Psychology:
Learning, Memory, and Cognition
2016, Vol. 42, No. 11, 1798–1813

0278-7393/16/$12.00 http://dx.doi.org/10.1037/xlm0000271

Processing of Irregular Polysemes in Sentence Reading

Andreas Brocher
University at Buffalo, The State University of New York

Stephani Foraker
SUNY Buffalo State

Jean-Pierre Koenig
University at Buffalo, The State University of New York

The degree to which meanings are related in memory affects ambiguous word processing. We examined irregular polysemes, which have related senses based on similar or shared features rather than a relational rule, like regular polysemy. We tested to what degree the related meanings of irregular polysemes (*wire*) are represented with shared semantic information versus unshared information represented separately, like homonyms (*bank*). Monitoring eye fixations, we found that later context supporting the less frequent meaning of an irregular polyseme did not slow down reading compared with control conditions, whereas for homonyms it did. This indicates that in the absence of preceding biasing context, readers access a shared component of an irregular polyseme's representation. Additionally, when the same context words preceded the ambiguous word, both irregular polysemes and homonyms initially elicited longer reading times, but the observed reading slow-down was weaker and less persistent for irregular polysemes than homonyms, indicating less competition between meaning components. We interpret these results as evidence of a shared features representation for irregular polysemes, which additionally incorporates unshared portions of meaning that can compete. When preceding, biasing context is available, readers activate shared and unshared components of the senses, producing a more fully instantiated meaning.

Keywords: eye-movements, homonymy, lexical ambiguity, polysemy, semantic features

Supplemental materials: http://dx.doi.org/10.1037/xlm0000271.supp

Most words of a language have more than one interpretation. Yet, in most instances, such meaning ambiguity seems to not pose any serious problems in language comprehension. It is, therefore, essential for any theory of comprehension to determine how ambiguity is resolved, including how this may vary for different types of ambiguous words. Homonyms, such as *bat*, have unrelated meanings, whereas polysemes, such as *book* or *wire*, have multiple semantically related senses. Interestingly, homonyms have been investigated more systematically and extensively than polysemes, despite the fact that homonyms make up only a small portion of ambiguous words, whereas polysemes are far more ubiquitous.

Over the past decade or two, a debate about how polysemes are represented in the mental lexicon and processed in the presence and absence of context has developed (Beretta, Fiorentino, & Poeppel, 2005; Eddington & Tokowicz, 2015; Foraker & Murphy, 2012; Frazier & Rayner, 1990; Frisson, 2009, 2015; Frisson & Pickering, 1999; Klein & Murphy, 2001, 2002; Klepousniotou, 2002; Klepousniotou, Pike, Steinhauer, & Gracco, 2012; Klepousniotou, Titone, & Romero, 2008; Pickering & Frisson, 2001; Rabagliati & Snedeker, 2013; Rabagliati, Marcus, & Pylkkänen, 2010, 2011; Rodd, Gaskell, & Marslen-Wilson, 2002, 2004; Srinivasan & Snedeker, 2011; Williams, 1992). Factors such as type of polysemy, sense dominance, degree of sense relatedness, degree of semantic overlap, and biasing context are being investigated, but to date not thoroughly enough to reach clear conclusions. Here, we examine one type of polysemy—irregular polysemy—and provide evidence for an updated model of how readers represent and compute the appropriate sense of an irregular polyseme in sentence context.

Homonyms in Sentence Contexts

In investigations of polysemy representation and processing, a common benchmark comparison has been between homonyms and polysemes (Frazier & Rayner, 1990; Klein & Murphy, 2001; Klepousniotou, 2002; Klepousniotou & Baum, 2007; Klepousniotou et al., 2012; Pickering & Frisson, 2001; Rabagliati & Snedeker, 2013; Rodd et al., 2002). Hence, we first provide a brief summary of key effects found for homonyms in sentence contexts, which bear on how homonym meanings are represented.

Biased homonyms, which have two meanings of unequal frequency, reliably produce a *dominance effect*, where the more

This article was published Online First April 11, 2016.

Andreas Brocher, Department of Linguistics, University at Buffalo, The State University of New York; Stephani Foraker, Department of Psychology, SUNY Buffalo State; Jean-Pierre Koenig, Department of Linguistics, University at Buffalo, The State University of New York.

This experiment was part of Andreas Brocher's dissertation, and was presented to the audiences of AMLaP 2012 and the Mental Lexicon 2012. We thank the members of these audiences for fruitful discussions as well as Gail Mauner for valuable comments on earlier versions of this paper. We thank Carla Kuhl, Julie Yeh, and Kate Donelson for their assistance with data collection.

Correspondence concerning this article should be addressed to Andreas Brocher, who is now at Universität zu Köln, Institut für Deutsche Sprache und Literatur I, Albertus Magnus Platz, 50923 Köln, Germany. E-mail: abrocher@uni-koeln.de

frequent, dominant meaning is accessed more easily and quickly than the less frequent, subordinate meaning. For example, Duffy, Morris, and Rayner (1988) measured readers' eye movements as they read sentences that contained a homonym (*port* – harbor; wine) or an unambiguous control word (*soup*). The target word was followed by context that supported the subordinate meaning of the homonym: *Last night the port (soup) was a great success when she finally served it to her guests*. Duffy and colleagues found that reading times on the context region (*when she finally served it to her guests*) were longer following the homonym compared to the unambiguous control condition. This indicated that readers access the homonym's dominant meaning, and make an *immediate semantic commitment* (Frazier & Rayner, 1990) to that particular meaning. Readers then start integrating that meaning into the sentence, but when they later encounter the inconsistent subordinate context they need to adjust and reanalyze the interpretation of the homonym and sentence. Such dominance effects due to differential meaning frequency are taken to indicate that the multiple meanings of homonyms are stored separately.

Duffy et al. (1988) also showed that when context supporting the subordinate meaning preceded the homonym (the two clauses were reversed), readers took longer to read the later homonym than its control word. This *subordinate bias effect* on the homonym indicated that readers (automatically) access the dominant meaning of a homonym, which competes with the contextually boosted subordinate meaning, leading to slower processing at the homonym. Further research also shows that the degree to which the subordinate meaning is available for a homonym is modulated by the preceding sentence context (Dopkins, Morris, & Rayner, 1992; Simpson & Krueger, 1991; Vu, Kellas, & Paul, 1998; see also Binder & Rayner, 1998, 1999; Kellas & Vu, 1999; Martin, Vu, Kellas, & Metcalf, 1999; Rayner, Binder, & Duffy, 1999) and preceding discourse context (Colbert-Getz & Cook, 2013; Martin et al., 1999).

Therefore, a particular meaning can be activated through direct lexical access or contextual priming, and meanings of unequal frequency can compete or slow down processing, consistent with separately stored meaning representations. Overall, dominance effects for biased homonyms have been observed in numerous studies, and underline the robustness of the claim that homonym meanings have separately stored representations (Binder & Morris, 1995, 2011; Dopkins et al., 1992; Duffy et al., 1988; Folk & Morris, 2003; Kambe, Rayner, & Duffy, 2001; Rayner & Duffy, 1986; Rayner, Pacht, & Duffy, 1994; Sereno, O'Donnell, & Rayner, 2006; Simpson, 1981; Simpson & Burgess, 1985; Simpson & Krueger, 1991; Vu et al., 1998).

Several experiments have now compared polysemes with homonyms, using sentence reading measures (Frazier & Rayner, 1990; Pickering & Frisson, 2001), lexical processing tasks (Klein & Murphy, 2001; Klepousniotou, 2002; Klepousniotou & Baum, 2007; Klepousniotou et al., 2008; Rodd et al., 2002), imaging measures (Beretta et al., 2005; Klepousniotou et al., 2012; MacGregor, Bouwsema, Klepousniotou, 2015; Pylkkänen, Llinas, & Murphy, 2006), and computational modeling (Armstrong & Plaut, 2008, 2011; Rodd et al., 2004, see also Hino, Pexman & Lupker, 2006). Most studies have shown that polysemes are processed differently than homonyms, but not all (cf. Klein & Murphy, 2001; Rabagliati & Snedeker, 2013). A key factor, we observe, is that polysemy is not homogenous. Hence, different kinds of sense relatedness must be considered to fully explain existing experimental results and develop a comprehensive understanding of ambiguity resolution.

Types of Polysemy

The majority of reading experiments have focused on regular polysemy (Frazier & Rayner, 1990; Frisson, 2015; Frisson & Frazier, 2005; Frisson & Pickering, 1999; Pickering & Frisson, 2001). Regular polysemy picks out a noteworthy relation between senses that occurs in predictable, consistent contexts (Nunberg, 1995). A typical example is the metonymic relation between two senses of a word (e.g., *hospital*), where one sense is literal or concrete (*The hospital is centrally located*) and the other more abstract or figurative (*The hospital complained about a shortage of staff*). The metonymy, here, is that a complaining hospital refers to the people in it. The systematicity of this kind of metonymy (it applies to an entire class of concepts) can be modeled as a productive rule or referring function, such as person-for-institution, that allows one sense to be derived from the other default or base sense (Apresjan, 1974; Copestake & Briscoe, 1995; Cruse, 1986; Lakoff, 1987; Rabagliati et al., 2010, 2011). Note that there is often a directional restriction on the rule: hence, *complaining hospital* is an acceptable rule-based sense extension (as it involves the metonymy person-for-institution) and can be productively applied to other institution concepts such as *smart school* or *rich church*. On the other hand, *brick nurse* is not an acceptable rule-based derivation, as it would involve the converse metonymy institution-for-person, nor is *five-story teacher* or *Romanesque priest*.

Irregular polysemes, such as *wire*, have senses that are related in a more idiosyncratic way, and are not derived from one another via a productive rule (Apresjan, 1974; Eddington & Tokowicz, 2015; Rabagliati & Snedeker, 2013). These words, with less regular, unpredictable senses, are fairly common (Lehrer, 1990). The main claim we pursue in this paper is that a critical basis for the relation between irregular polyseme senses is sense overlap, in the form of particular semantic features. For example, *wire* has the two senses of flexible filament and listening device. The overlapping semantic content would be the shared features of, for example, *metal*, *cylindrical*, *thin*, or *small*, and could provide salient or diagnostic features for this particular ad hoc relation. Semantic feature overlap could also apply to metaphoric polysemes, a subclass of irregular polysemy. Semantic features that serve as a basis of relation between senses induce a similarity relation (Gentner, Bowdle, Wolff, & Boronat, 2001; Giora, 1999; Glucksberg, 2003; Glucksberg & Keysar, 1990; Kintsch, 2000). Provided that the metaphoric reading is sufficiently frequent, it could include such shared features as part of its stored representation.

Current Models of Polysemy Representation

How a polyseme's senses are related or overlap could well have effects on how the different types of polysemes are represented in the mental lexicon. On the one hand, when senses are related in a lexical or conceptual way, one might assume that the senses share some amount of meaning. We will group variants of this approach under *shared sense representations*, but note that there is variability in what "shared" captures. Most variants fit the systematicity of regular polysemy sense extensions better, but do not fare as well for the more idiosyncratic relations of irregular polysemy. A second possibility for polysemes is *separate sense representations*. This approach proposes that different senses are listed separately in the mental lexicon, which captures some aspects of irregular polysemy well (some features are not shared), but overlooks semantic content that can be shared. We therefore propose a new model of representation and processing for irregular polysemy that can explain some of the conflicting evidence in the literature, and we report a first experiment that tests this model.

Models of polyseme processing and representation that are compatible with a common or shared component emphasize the relatedness or similarity of senses. By and large, variants of this approach were developed for regular polysemy and capture the productive nature of such sense relations. The most discussed variant for sentence contexts involves an *underspecified* representation (Frisson, 2009, 2015; Frisson & Pickering, 2001; Pickering & Frisson, 2001; see also Frisson & Pickering, 1999). An underspecified representation "encompasses all semantically related interpretations of a word known to the reader" (Frisson, 2009, p. 116), and is the same for all those senses.

Unlike for homonyms, the comprehender activates just the common, underspecified content of a polyseme during lexical retrieval, making only a *partial semantic commitment* (Frazier & Rayner, 1990). Unless it is grammatically required to commit to one sense (e.g., subject-as-agent principle, Fishbein & Harris, 2014), the underspecified portion is sufficient to access the polyseme, and then readers can fill out or home-in on the intended sense more fully, depending on several factors (Frisson, 2009), including the presence of later disambiguating context, or a choice point, like the end of a sentence. Importantly, the underspecification model posits that sense frequency does not affect polyseme retrieval per se (e.g., Pickering & Frisson, 2001).

In an early investigation, Frazier and Rayner (1990) examined primarily regular polysemes (e.g., *newspaper*, *novel*, *poem*): *Unfortunately the newspaper was destroyed, lying in the rain/managing advertising so poorly*. Earlier and later eye fixation measures indicated that following a polyseme, readers did not show any processing cost for the subordinate sense (*managing advertising*). Following a homonym, though, readers did show the expected slow-down on contexts that supported the subordinate meaning: *Everyone thought the ring infuriated Susan, sounding so shrill/looking so cheap*. Additionally, when the subordinate-bias context preceded the ambiguous word, readers took significantly longer to read homonyms than polysemes (and controls), for early and later reading measures. Critically, there was no subordinate bias effect for the polysemes. In sum, the results are consistent with an initial underspecified representation common to both senses, and later fleshing out of the subordinate sense without noticeable difficulty.

Frisson and colleagues have investigated regular polysemy that involves metonymy relations, such as building-institution (*school*), place-event (*Vietnam*), or producer-product (*Dickens*). In one study (Frisson & Pickering, 1999), they compared a metonymic noun's literal versus established metonymic sense, finding that it was equally easy to process the two senses. This was particularly so for early measures of processing, with some indication of integration slow-down in later measures for the metonymic sense (see also Frisson & Pickering, 2007; Pickering & Frisson, 2001). Furthermore, in a follow-up analysis, they found that sense frequency did not account for regular polyseme processing times, contra any kind of dominance effect. For regular polysemy, then, a single, underspecified representation that is not affected by sense frequency is well supported. However, this approach does face the difficulty of defining what the underspecified content actually *is*.

An alternative approach to regular polyseme representation, still falling within shared sense representations, is that the systematic relations between polyseme senses are represented in generative structures that allow meaning to shift according to the encoded relations. Sense generation can occur at a conceptual level, where the underspecified representation first activated is a base sense, plus using a relatedness rule, such as a count-mass or animal-meat relation for *chicken*, *ostrich*, or even *penguin* (Copestake & Briscoe, 1995; see also Lakoff, 1987; Langacker, 1987; Lehrer, 1990; Rice, 1992; Tuggy, 1993). There are some constraints on this approach from experimental investigations. First, it is not the case that a literal, concrete sense is accessed first, and a metonymic, figurative sense after. Second, the base sense is not always the dominant one. Frisson and Pickering (1999) found no reading time differences between literal versus figurative senses for familiar place-for-institution (. . . *walked/talked to the school*) or place-for-event metonymies (. . . *hitchhiked around/protested during Vietnam*), and that sense dominance did not correlate with reading times.

Sense generation may also occur at the lexical level. Pustejovsky (1995), for example, posits a rich lexical structure that allows the online computation of any sense from an underspecified base. There is substantial psycholinguistic evidence of semantic coercion or type-shifting costs associated with regular polysemy (e.g., *book* as an entity → event: *the gentleman started Dickens*) supporting this approach (Frisson, Pickering, & McElree, 2011; Pickering, McElree, Frisson, Chen, & Traxler, 2006; Pylkkänen, Martin, McElree, & Smart, 2009; Pylkkänen & McElree, 2006; Traxler, Pickering, & McElree, 2002).

Lastly, still falling within shared sense representations is a core meaning variant. A core meaning is said to contain just the common features that overlap between senses (Nunberg, 1979, 1995, see also Williams, 1992), and any more specific interpretation is left vague (Ruhl, 1989). A recent take on a core representation for regular polysemy is proposed by Klepousniotou et al. (2008; see also Klepousniotou, 2002; Klepousniotou & Baum, 2007; Klepousniotou et al., 2012). Focusing on degree of sense relatedness, Klepousniotou et al. (2008) classified items into high relatedness (metonymic polysemy), moderate relatedness (a mix of metonymic and metaphoric polysemy and homonyms), and low relatedness (homonyms) groups. Using a sense priming task, they found a small cost for metonymic (i.e., regular) polysemy when switching from the dominant to subordinate sense, but no cost in the other direction. They concluded that metonymic/regular polysemes therefore have a unitary lexical representation with a core meaning that generally maps to the dominant sense, and which is always activated, irrespective of context. They proposed that the core representation would include "only those semantic features that are in common and compatible across all possible senses of the word" (Klepousniotou et al., 2008, p. 1538).

Yet, such a core meaning approach has been argued against, primarily for regular polysemy (Klein & Murphy, 2001, 2002; Murphy, 2007; Rice, 1992; Taylor, 2003), because it is difficult to identify what the semantic properties in common would be between metonymic senses. Using the example of *rabbit*, if one takes the live animal sense as the dominant one, features such as + animate, + farm animal, + furry, + hop, and + big ears are plausible (features taken from Klepousniotou et al., 2008, p. 1538). But, we point out that the subordinate meat sense does not really have any of these features in common, and some are not compatible (+ animate, + hop) or are typically incompatible (+ furry, + big ears). In the same way, the meat sense could have the feature + edible, but this feature is unlikely to apply to the live furry animal sense. Hence, although there is a regular relation between the two senses, it does not necessarily entail that they share the same physical features, the same functions, and so on. We would like to note here, though, that a core meaning approach can likely work well for irregular polysemy, capturing features in common between senses that are related more idiosyncratically. We develop this idea further as part of our model, below.

Turning to the two other conditions in Klepousniotou et al. (2008), they found switching costs consistent with dominance and subordinate bias effects for low-overlap homonyms and moderate-overlap polysemy (including irregular polysemy). These results replicated Klein and Murphy (2001), discussed below (and whose items they incorporated), producing the conclusion that metaphoric polysemes with moderate relatedness between the senses have distinct, independent lexical representations, like those of homonyms.

This introduces the major contrast to shared sense representations, which is *separate sense representations*, positing that representations for polysemes are largely organized the same way as those for homonyms. This account can be described as a Sense Enumerated Lexicon (Clark & Gerrig, 1983; Pustejovsky, 1995) or a List Model of word representation (Srinivasan & Snedeker, 2011). One major drawback of this approach, though, is that the separate senses cannot easily support generativity—shifting from one sense to another. This is a problem especially for regular polysemy (discussed in Klepousniotou, 2002; Srinivasan & Snedeker, 2011), but because irregular polysemy is more idiosyncratic and typically nongenerative, separate sense representations is a viable alternative.

Early comprehension evidence for separately stored sense representations comes from Klein and Murphy (2001, 2002). In the 2001 investigation, they primed consistent (*wrapping paper—shredded paper*) or inconsistent (*wrapping paper—liberal paper*) senses of a polyseme (*paper*) during lexical processing tasks, and found that consistent sense pairs produced facilitation, and inconsistent pairs produced inhibition. Polysemes were also processed quite similarly to homonyms, providing evidence favoring separate representations for polysemes. As pointed out by later investigators, though, their items included both regular and irregular polysemes (Foraker & Murphy, 2012; Klepousniotou et al., 2008), which makes it difficult to ascertain to what degree the type of polysemy and basis of sense relatedness played a role in their results.

Applying the Klein and Murphy (2001) materials to sentence contexts, Foraker and Murphy (2012) further assessed evidence for separate sense representations, testing for dominance effects like those found for separate homonym meanings. The polysemes were again a mix of regular and irregular polysemes, and they varied in both bias (highly biased to balanced) and sense similarity (rated post hoc). The authors reported an overall dominance effect, where (a) following neutral context, the dominant sense was accessed more easily than the subordinate sense, (b) following consistent context, the dominant sense was accessed more easily than the subordinate sense, and (c) following inconsistent context, there was greater between-sense competition when switching from dominant-to-subordinate sense than the other direction—all consistent with the patterns typical of homonyms. Based on such dominance effects, which had not been found for polysemes in sentence contexts before, they argued for separate sense representations.

We note, however, that no direct comparison to homonyms was made in Foraker and Murphy's (2012) sentence reading experiments. Additionally, dominance effects for polysemes did not occur in early reading measures, nor on the polyseme itself, which would be expected if polysemes have separate entries like homonyms (cf. Duffy et al., 1988). Furthermore, sense similarity modulated the dominance effects, such that the more similar two senses were, the smaller the dominance effect was (see also Klepousniotou et al., 2008). We suggest that these results are partially consistent with a shared representation account, considering that polysemes with greater similarity between senses could share more semantic information, and consequently produce no dominance effect (with no preceding biasing context, Frazier & Rayner, 1990; Frisson & Pickering, 1999).

Recently, using a production task, Rabagliati and Snedeker (2013) compared regular (*corn* – stalk of corn; kernels of corn) and irregular polysemes (*button* – shirt button; emergency button) in a picture-labeling task. They found that adult participants produced more ambiguous bare noun labels for irregular polysemes (*button*) at levels similar to homophones (*bat*), whereas regular polysemes led to statistically fewer ambiguous labels (*stalk of corn*). They argued that irregular polysemes therefore have separate sense representations, whereas regular polysemes have a core representation common to both senses (see also Ferreira, Slevc, & Rogers, 2005).

In sum, investigations of regular polysemy in sentence contexts tend to support some kind of shared representation. Several experiments have found no dominance effects or subordinate-bias effects, consistent with an underspecified representation. Although it is unclear what the content of an underspecified or core meaning representation for regular polysemy may be, a key conceptual or lexical relation, perhaps in the form of a generative rule, may be sufficient. Additionally, investigations that incorporate irregular polysemy (sometimes distinguished from regular polysemy, sometimes not), tend to support some kind of separate representation for the senses, and show dominance effects linked to sense frequency. Yet, irregular polysemy incorporates some semantic overlap in a way that is different from regular polysemy.

Toward a Shared Features Model

Here, we propose a shared features model for irregular polysemy, where the representations of the two senses are divided into shared and unshared feature components for each sense. The shared features portion consists of semantic features in common, and functionally is similar to the idea of semantic overlap inferred in an underspecified representation. The shared features portion is also similar to a core meaning shared between senses, although we do not claim that the shared portion carries sense frequency information. We would not claim that the core meaning is essentially the dominant sense (cf. Klepousniotou et al., 2008), but rather it is a subset of features common to and compatible with the two senses at issue.

Going beyond the underspecified and core meaning approaches, we also propose that the involvement of unshared features and their associated frequencies are critical for full sense commitment of irregular polysemy. The role of unshared semantic features has been less well articulated in previous research and is a novel component in our model. Unshared features do not overlap among the senses, but as a particular sense is filled out beyond the shared portion, (appropriate) unshared features become activated, too. The filling out could proceed relatively automatically, as with a spreading activation mechanism (e.g., Rodd et al., 2004), or could be instigated by supporting context.

An important aspect of the proposed model is that unshared features carry sense frequency information. For a biased polyseme, there would be a shared features component in common, and then dominant-sense-related unshared features for the one sense, and subordinate-sense-related unshared features for the other. The key difference between a shared features model and underspecification is that dominance effects can occur during polyseme access when unshared feature components of a sense are activated. Additionally, when dominance effects do occur, they should (a) vary in strength according to the biased sense frequencies, and (b) overall, be weaker than those found for homonyms, because the dominance effects are driven by the unshared portions, which are smaller for irregular polyseme senses than the wholly unshared homonym meanings.

The Current Experiment and Predictions

Because of the mixture of polysemy types and variability in sense relatedness across the existing experiments to date, it is still not fully clear how polyseme senses are represented, and why sentence processing differences occur. Our first question, then, addressed whether biased irregular polysemes are better characterized by separate or shared feature representations of a polyseme's senses. We compared reading times for irregular polyseme conditions with those for homonym conditions. If irregular polysemes pattern with homonyms in sentence reading, we can conclude that both kinds of ambiguous words are likely to be represented similarly, and that the multiple readings are represented separately for both kinds of ambiguity. On the other hand, different reading patterns would suggest that homonyms and irregular polysemes are represented differently. In particular, our shared features model predicts the absence of dominance effects for polysemes when a later disambiguating context region supports the subordinate reading. To test this, we included sentences in which the ambiguous word appeared in the first clause, and context supporting the subordinate meaning/sense appeared in the second clause. This prediction follows from the assumption that a (partial) shared representation allows readers to delay commitment to one sense, and to later fill out the interpretation without a costly reanalysis attributable to retrieval and integration of the competing sense in the preceding clause, which occurs for biased homonyms. Note, also, that in these conditions, the underspecification account and shared features model predict similar processing outcomes, albeit based on different assumptions about what constitutes a shared representation.

Our second question tackled the unshared portions of sense representation for irregular polysemes. One important characteristic of the underspecification account is that an underspecified node is always accessed first when the polyseme is encountered. Then, this node is filled out to home-in on a full interpretation when informative context is available. Note, however, that information about the senses' frequencies is not associated with the underspecified node, which leads to the explicit prediction of an absence of dominance effects when the polyseme is accessed and a sense selected, whether there is (subordinate-bias) context present or not (Pickering & Frisson, 2001).

For regular polysemy, such "blindness" to sense frequency may be plausible. Because an underspecified representation encompasses all senses known to the reader, yet is the same for all of those senses (Frisson, 2009), it may not contain much information that could differentiate one sense from the other. In contrast, for irregular polysemy, it is likely that there are more semantic features shared between senses, which, in the absence of a predictable relation, could be the basis for sense overlap, as in the *wire* example, above. The degree of overlap versus nonoverlap between senses should drive ease of settling on one sense rather than another when later context is processed, and contribute to how easily the comprehender can switch from one sense to another.

To be able to distinguish the underspecification account from our shared features model, we also included conditions with preceding biasing context. Here, the underspecification model predicts that polysemes are accessed like unambiguous controls and therefore unlike homonyms. This is predicted because disambiguating context does not affect sense retrieval or selection but, if at all, only later stages of sentence interpretation (Frisson & Pickering, 1999; Pickering & Frisson, 2001). In contrast, the shared features model predicts that readers make use of the contextual information to guide lexical retrieval and sense selection. Put differently, readers should retrieve not just the shared representation portion but, additionally, be biased to make a choice as to which sense to more fully access. Importantly, if the preceding context is consistent with an irregular polyseme's subordinate sense, when one then reads the polyseme, the unshared features unique to that subordinate sense could be contextually boosted, while unshared features unique to the word's dominant sense would become activated through sense frequency bias (on the polyseme). Because the unshared features of both senses would then be activated, competition should arise between the senses that would slow down the sense resolution process, resulting in longer reading times for irregular polysemes than control words in early measures (rather than in late measures that reflect potential integration difficulties).

In other words, in the presence of preceding subordinate-bias context, the lexical access of biased irregular polysemes should show some similarity to that of biased homonyms. Critically, the magnitude and extent should not be the same. The two senses of an irregular polyseme, unlike the two meanings of a homonym, share features that support sense resolution. The two meanings of homonyms are incompatible, but the two senses of an irregular polyseme are only partially incompatible. Thus, our second prediction was that with preceding subordinate-bias context, readers will show a subordinate bias effect on the polyseme (or soon after), but have less difficulty resolving biased polysemes to their subordinate sense than for biased homonyms to their subordinate meaning.

Method

Participants

Eighty undergraduate students from Buffalo State College (52 females, average age of 22 years) participated for course credit. All participants were native speakers of American English, schooled in the U.S., had no history of a language or learning disability, and reported normal or corrected-to-normal vision. Six participants were also native bilinguals who reported that English was their dominant language for speaking on a daily basis, and for reading and academics; these participants were distributed across the counterbalancing stimuli lists. Data from an additional four participants were excluded because their comprehension accuracy was below 80%.

Materials & Design

The 18 biased homonyms and 18 biased irregular polysemes were normed for meaning/sense similarity, meaning/sense dominance, and the familiarity of ambiguous words' less frequent meanings/senses. Norming was done across a number of cycles. Meaning similarity was defined as the degree to which speakers judge the two interpretations of an ambiguous word to be semantically similar based on physical, functional, or other properties. Meaning/sense dominance was defined as the relative frequency of two different interpretations of an ambiguous word. A summary of norming results can be found in Table 1. More specific norming details are provided in the Online Supplement.

Similarity norming. We performed several cycles of similarity norming. In each one, 20 monolingual native speakers of American English (at the University at Buffalo), none of whom participated in the main experiment, were presented with booklets containing pairs of single sentences. Each sentence pair had one word in common, see (1).

(1) (a) Paul wanted to deposit all his cash but the *bank* was already closed.

(b) The couple went for a nice, long walk alongside the *bank*.

Participants were instructed to judge the semantic similarity between the two tokens of the underlined word. They were asked to base their judgments on the following questions: *Can the two meanings appear in similar contexts? Do they share physical or functional properties? Do they taste, smell, sound, or feel similarly? Do they behave similarly?* These questions were provided to help participants base their judgments on semantic features rather than noninterpretable ad hoc associations (like *go there every week* or *rhymes with tank* for the homonym *bank*). Participants were instructed to provide a similarity score ranging from 1 for *not similar at all* to 7 for *the very same meaning*, and encouraged to use the whole range of the scale. Each participant provided similarity scores for 50 sentence pairs that included 16 homonyms (*bank*), 16 polysemes (*wire*), and 18 neutral filler words (*origami*).

The final 18 homonyms that were selected for the sentence reading experiment had a mean similarity score of 1.29 ($SD = 0.17$), and the 18 polysemes had a mean similarity score of 3.16 ($SD = 0.65$), a significant difference as indicated by an unpaired t test (two-tailed), $t(34) = 12.0$, $p < .0001$. Neutral filler words had a mean similarity score of 6.46 ($SD = 0.49$), indicating that the irregular polysemes selected for the main experiment fall between the homonyms and unambiguous words.

Dominance norming. In each dominance norming cycle, 20 monolingual native speakers of American English (at the University at Buffalo) who did not participate in the main experiment were presented with booklets consisting of words, one per line. Every word was presented five times in a row and was followed by a blank line. Participants were instructed to write down whatever came to their minds, as single words, phrases, or entire sentences. Each booklet consisted of 48 different words (i.e., participants provided 240 associations in total), of which eight were homonyms (*bank*), eight polysemes (*wire*), and 32 neutral filler words (*origami*).

Two different raters (native speakers of American English who received training in the task) decided for each produced association whether it belonged to one of the ambiguous word's targeted meanings (e.g., wire-cable or wire-spy), to a different or noncomprehensible meaning (e.g., *bang*), or either meaning (e.g., *beautiful*). Raters were instructed to only assign a particular association to one of the targeted categories (e.g., *cable*) when it could not also be assigned to the competing category (*spy*), even when the association was more related to one than the other. Disagreements were resolved by discussion. A particular association was assigned to the category *different* when no agreement was reached. After resolution, overall agreement was above 90%.

Table 1
Norming Specifics of Materials

Ambiguity	Dominance	Similarity	Familiarity	Plausibility	Frequency_{WF}	Frequency_{CD}	Letters	Syllables
Homonyms	.87 (.02)	1.28 (.04)	6.13 (.19)	4.59 (.16)	37.74 (7.48)	10.84 (2.49)	4.50 (1.03)	1.33 (.31)
Controls	—	—	—	5.04 (.17)	65.58 (28.57)	12.34 (2.82)	4.44 (1.02)	1.22 (.28)
Polyseme	.89 (.02)	3.16 (.15)	6.43 (.24)	4.73 (.14)	71.66 (24.07)	18.44 (4.23)	4.72 (1.08)	1.28 (.29)
Controls	—	—	—	5.09 (.16)	149.50 (72.40)	20.60 (4.73)	4.78 (1.10)	1.28 (.29)

Note. Dominance of meanings/senses (.5 = *not biased* to 1 = *strongly biased*), Similarity of meanings/senses (1 = *not similar* to 7 = *same meaning*), Familiarity of subordinate meanings/senses (1 = *not familiar* to 7 = *very familiar*), and Plausibility of experimental sentences (1 = *not plausible* to 7 = *very plausible*) were obtained in local norming studies; $Frequency_{WF}$ = raw lexical frequencies for ambiguous and control words from Subtlex (per 1 million); $Frequency_{CD}$ = percentage of films in which the ambiguous or control word appears in Subtlex (of 8,338 films); *Letters* = number of letters of ambiguous and control words; *Syllables* = Number of syllables of ambiguous and control words; Standard errors are presented in parentheses.

For all selected items, we chose the meaning that had been produced most often as the dominant interpretation and calculated the dominance score relative to the other, subordinate meaning. We only considered the two intended readings in computing the reported dominance scores so that the proportions of the dominant and subordinate readings always summed to 1. We used this type of dominance score because it is less susceptible to noise from unresolved raters' disagreements or noncomprehensible productions than if responses assigned to the *different* category had been included. The 18 homonyms and polysemes that we selected for the main experiment were carefully equated, with mean dominance scores of .87 (SD = .09) and .89 (SD = .07), respectively. An unpaired, two-tailed t test confirmed that the difference was not statistically significant, $t(38) = 0.42$, $p = .599$.

As a reviewer remarked, the n-size we used for dominance norming is smaller than some past research (e.g., Twilley, Dixon, Taylor, & Clark, 1994), which may render our dominance scores less reliable. We did choose to use local norms to better capture meaning variation in our target sample (e.g., Binder & Rayner, 1998), as opposed to a national or international sample, or the use of dictionary entries. However, our procedures did not differ remarkably from other past research on meaning or sense ambiguity (Duffy et al., 1988; Foraker & Murphy, 2012; Klepousniotou et al., 2008; Pickering & Frisson, 2001, see also Frisson, 2015). Additionally, note that we presented equal numbers of homonyms and polysemes to each participant, which should at least equalize the reliability of dominance estimates for polysemes and homonyms.

Sense familiarity norming. Additionally, 20 participants from Buffalo State College rated the familiarity of candidate ambiguous words' subordinate meaning/sense, to ensure that readers could access the less frequent interpretation. Participants rated the critical word (which was underlined in a sentence) on a scale from 1 for *completely unfamiliar* to 7 for *completely familiar*, and then provided "a short phrase or definition of the underlined word's meaning as it is expressed in the sentence [they] read." The difference in familiarity of subordinate homonym meanings (M = 6.13, SD = 1.1) and subordinate polyseme senses (M = 6.43, SD = 0.8) was not significant as shown by an unpaired t test (two-tailed), $t(34) = 1.64$, $p = .111$.

Experimental sentence stimuli. To assess the effects of biasing contextual information on the processing of biased irregular polysemes and homonyms, we followed the design of Duffy et al. (1988) and Frazier and Rayner (1990). Because of norming restrictions, it was impossible to find homonyms and irregular polysemes that would allow for a direct comparison of their reading times. We therefore constructed sentence pairs for both kinds of ambiguity. One sentence of each pair contained an ambiguous word and the other sentence a matched control word. To reduce the probability that properties of the control words affected reading time differences, we selected control words that (a) plausibly fit each sentence and (b) were matched as closely as possible in plausibility, frequency (using the contextual diversity measure in Subtlex; Brysbaert & New, 2009), number of letters, and number of syllables across conditions (homonyms vs. irregular polysemes, see Table 1). There were no significant differences within any set of ambiguous and unambiguous control words, as indicated by unpaired, two-tailed t tests, ts < 1.2, ps > .2.

For each ambiguous and control word pair, we constructed sentences with two clauses. Table 2 provides a set of example sentences. One clause always contained either a biased homonym (*bank*) or a biased irregular polyseme (*wire*) while the other contained context information that always disambiguated the ambiguous word toward its subordinate interpretation. We will refer to the sentences in which the ambiguous word appears in the first clause and the biasing context in the second clause as *Context After* sentences. We will refer to sentences in which the ambiguous word appears in the second clause and the biasing context in the first as *Context Before* sentences. To reduce the number of counterbalanced presentation lists, we constructed two carrier sentences per ambiguous/control word (versions A and B, see the Online Supplement). For a particular list, if one version of an experimental sentence had an ambiguous word in it, the other list had its control word. This pattern was counterbalanced across lists (see Duffy, Morris, & Rayner, 1988).

Sentence plausibility norming. We collected plausibility judgments for each sentence in the Context Before condition, from 46 native American English-speaking participants (at Buffalo State College). Sentences were presented in the Context Before condition to maximize the sense/meaning's contribution to the plausibility judgments. As well, the shared features and separate entries models make similar predictions for this condition (see above). The experimental sentences were interspersed among 36 plausible distractor sentences without noticeable ambiguity, and 18 less plausible ones, and then counterbalanced across two paper-and-pencil forms in two fixed pseudorandom orders. Participants were asked to rate sentences on a scale from 1 for *makes no sense at all* to 7 for *makes complete sense*. Sentences containing homonyms and their control words had plausibility scores of 4.59 (SD = 1.0) and 5.04 (SD = 1.0), respectively, and those with polysemes and their control words were 4.73 (SD = 0.9) and 5.09 (SD = 1.0), respectively.

A generalized linear regression model revealed that sentences with ambiguous words were rated as less plausible than sentences with respective control words, $\beta = -0.40$, $SE = 0.17$, $t = -2.41$, $p = .019$. It is important to note, however, that sentences containing homonyms and polysemes still fell within the range of plausible sentences, that is, on average, they approached the "makes

Table 2
Example Materials

Ambiguity	Context	Sentence
Homonym	After	Michael didn't like the |**bank** (lake) |in the |suburbs, because the |*fishing* |was not |very good.
Polyseme	After	Because the |**wire** (bomb) |was well |hidden, the skilled |*spy* |of the |agency remained undetected.
Homonym	Before	Because the *fishing* was not very good, Michael didn't like the |**bank** (lake) |in the |suburbs.
Polyseme	Before	The skilled *spy* of the agency remained undetected, because the |**wire** (bomb) |was well |hidden.

Note. For illustration purposes only, the ambiguous word appears in bold and its matched control follows in parentheses, the disambiguating word appears in italics, and the pipe symbols indicate analysis regions.

good sense" category of the norming experiment. In addition, and more importantly, the observed differences in plausibility did not significantly differ between homonyms and polysemes, $\beta = -0.10$, $SE = 0.34$, $t = -0.29$, $p = .776$. Thus, any difference between homonyms and polysemes cannot be explained by differences in sentence plausibility, nor by differences in their control conditions.

Counterbalancing. The full set of 144 experimental sentences was counterbalanced across four presentation lists. Each list contained 36 sentences with an ambiguous word (18 homonyms and 18 irregular polysemes) and 36 sentences containing a control word. An ambiguous sentence and its control sentence always appeared in different sentence versions on the same list (e.g., on List 1, *horns* appeared in the Version A sentence, and its control, *bones*, appeared in the Version B sentence). Importantly, both the "A" and "B" versions of a sentence appeared in the same context within a presentation list to reduce between-subjects variability for the ambiguous versus control word factor (e.g., *horns*-Version A and *bones*-Version B both appeared in the Context After condition). The experimental items were interspersed with 72 filler sentences designed to distract participants from the experimental manipulation, and presented in a unique random order for each participant. A Yes-No comprehension question followed 36 of the fillers (half "yes", half "no"), which were the same on all four lists, representing about 25% of the total trials.

Procedure & Analytic Plan

Participants were seated in front of an SR Research Eyelink 1000 eye-tracker configured with the tower mount. We used SR Research's Experiment Builder program template for reading experiments (version 1.10.1) to present stimuli and record eye-movements. Each sentence was displayed all at once on one line, left-justified, about 1/3 down the monitor, in 11-point Lucida Console fixed-width font. Participants' eyes were 55 cm from the screen, such that three letters subtended 1° of visual angle. They were instructed to read each sentence normally for comprehension and to press a "DONE" button on the response box when they felt they understood it. The next screen displayed either a comprehension question, answered with a "yes" or "no" button, or a READY? message, responded to with the "DONE" button. Comprehension and READY? questions appeared 2/3 down the screen in the same font as the sentences. About half way through the trials, participants took a 5-min break. Eye position was recalibrated before they completed the experiment. Prior to beginning the experiment, participants first completed five practice trials. The experiment took about 45 min.

Dependent measures. Six dependent measures are reported: *First fixation*, *single fixation*, *first pass*, *first pass regressions*, *regression path duration*, and *total reading time* (Rayner, 1998). *First Fixation* refers to the duration of the first fixation in a region. *Single fixation* refers to the duration of the first fixation in a region when only one fixation was made before exiting the region. *First pass* is defined as the sum of fixations in a region before exiting that region to the right or left. (For convenience, we use the term "first pass" for single word regions as well, rather than "gaze duration.") *First pass regression* refers to the percentage of regressions out of a region during a first pass reading. *Regression path duration* (also called *go-past time*) is the sum of all fixations in a region, including regressive fixations to earlier parts of the sentence, before progressing past the region's right boundary. *Total reading time* is the sum of all fixations in a region.

Researchers have suggested that early measures (*first fixation*, *single fixation*, *first pass*) reflect lexical access and are reliably influenced by lexical frequency, meaning/sense frequency, and previous context. We interpreted such early measures as reflecting aspects of lexical representation, or the first stage of accessing a separate meaning/sense versus a shared features (or underspecified) sense. Later measures (*first pass regression*, *regression path duration*, *total reading time*) tend to reflect different aspects of processing, such as postlexical integration and reanalysis (for reviews see Pickering, Frisson, McElree, & Traxler, 2004; Rayner, 1998). We interpreted the three later measures as reflecting meaning/sense integration, which could include meaning/sense competition, or a homing-in stage.

In addition, we checked that readers landed on the ambiguous word and disambiguating word in computing first pass reading time. If they did not, we expanded the ambiguous and disambiguating word regions to include four characters to the left to allow for parafoveal preview (Blanchard, Pollatsek, & Rayner, 1989, see Rayner, 1998 for a review), which in most cases included a determiner (*the notes*) or short modifier (*new belt*). If there was still no first pass time (4.0% in the Context After and 8.0% in the Context Before conditions), the trial was excluded for all measures. For spillover regions, following either the ambiguous or disambiguating word, we never expanded the word regions of interest.

Regions of analysis. We recorded and analyzed eye movements for two sentence regions in the Context Before sentences and two different regions in the Context After sentences, illustrated in Table 2. For Context After conditions, the *Context words* region consisted of the first words within sentences that were strongly biased toward an ambiguous word's subordinate interpretation (e.g., *spy* for *wire*). The two subsequent words (spanning 6–8 characters, e.g., *of the*) served as the *Context spillover* region. These areas were of most interest in the Context After condition because readers had already accessed the ambiguous word in the preceding clause and could more fully commit to or revise their initial interpretation in the second clause. In the Context Before conditions, the regions of interest were the *Ambiguous word* itself (e.g., *wire*) and the *Ambiguous word spillover* region (spanning 5–7 characters, e.g., *was well*). In the Context Before conditions, reading times on context regions were not of interest and will not be discussed.[1]

Results

Statistical analyses were conducted in R (R Core Team, 2013). We performed linear mixed effects regression models for each reading measure (dependent variable) and each region of interest separately, using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). Cohen's *d*s were calculated using the lsr package (Navarro, 2014). The predictors of Ambiguity (homonym vs. polyseme) and Word Type (ambiguous vs. control word) were included

[1] As expected, mixed model regression analyses for disambiguating words in the Context Before conditions revealed no significant reading time differences, $ts < 1.2$, $ps > .2$.

Table 3
Context After Conditions: Dependent Measures and Effect Sizes

		Context words			Context words spillover		
Measure	Ambiguity	Ambiguous	Control	*d*	Ambiguous	Control	*d*
First fixation	Homonym	250 (3)	251 (5)	.03	245 (4)	237 (4)	.23
	Polyseme	249 (4)	251 (3)	.06	245 (4)	247 (4)	.04
Single fixation	Homonym	249 (4)	250 (6)	.01	247 (5)	239 (4)	.21
	Polyseme	254 (4)	251 (5)	.11	249 (5)	250 (4)	.01
First pass	Homonym	292 (5)	295 (5)	.01	303 (7)	282 (6)	.30
	Polyseme	297 (6)	300 (5)	.05	294 (7)	295 (6)	.00
First pass regression	Homonym	17.6 (4.3)	16.1 (4.1)	.10	14.0 (3.7)	11.5 (3.6)	.09
	Polyseme	20.2 (4.5)	18.9 (4.4)	.07	10.3 (3.4)	10.8 (3.5)	.03
Regression path	Homonym	375 (11)	361 (9)	.07	371 (12)	336 (10)	.28
	Polyseme	392 (10)	378 (9)	.09	339 (10)	341 (9)	.03
Total time	Homonym	387 (10)	373 (9)	.11	391 (11)	364 (9)	.14
	Polyseme	386 (9)	379 (9)	.07	375 (10)	378 (9)	.05

Note. Standard error appears in parentheses following the mean, both in ms. First pass regressions are presented as proportions. d = Cohen's d.

as fixed effects, and sum-coded before analysis. Each regression model included the interaction of ambiguity and word type, and random intercepts for participants and items. Following Barr, Levy, Scheepers, and Tily (2013), random slopes for participants and items were kept maximal and included the Ambiguity × Word type interaction. In case a model failed to converge, the interaction term was removed and only main effects were included in the random slopes. We determined p values on the assumption that, with many observations, the t distribution converges to the z distribution. One may also use the decision rule that an absolute t value of 2 indicates a significant effect (Gelman & Hill, 2007). Follow-up t tests are reported for comparisons by participant and item means.

In preparing data for statistical analyses, we excluded fixations shorter than 60 ms and longer than 1000 ms (6.2% of the data). If two fixations were within three characters of each other, and one of the fixations was shorter than 60 ms, they were merged into a single fixation. (Analyses with an 80 ms cutoff produced similar results). All reading times included in the analyses were log-transformed using Box-Cox power transformations to reduce skewness of the data. Overall comprehension question accuracy was 87%.

Context After Conditions

Dependent measures and Cohen's ds are shown in Table 3. Results of the regression models can be found in Tables 4 (Context word) and 5 (Context word spillover). Because word properties that affect reading times were not controlled across the factor of ambiguity, that is, between homonyms/controls and polysemes/controls, main effects of ambiguity are not meaningful and will therefore not be discussed. Before reviewing reading times of subordinate-bias context words, we should point out that no significant reading time differences were found on the homonyms or polysemes in this condition, ts < 1, ps $> .3$, replicating previous research (e.g., Duffy et al., 1988).

Turning to the disambiguating context words, inspection of Table 3 shows that subordinate-bias contexts were not read more slowly following ambiguous words compared with matched control words. This was true for homonyms and polysemes. Consequently, there were no statistically reliable effects on the context word (see Table 4). These data suggest that readers did not experience a conflict between mutually activated meanings or senses when they encountered the first disambiguating word.

Reading times on the following context spillover region indicated that readers did slow down shortly after the first disambiguating word. A significant interaction for the early measure of first pass times occurred, and the later measure of regression path (see Table 5). Crucially, longer reading times were restricted to homonyms, in both earlier measures (Table 3 Cohen's d, first fixation, single fixation, first pass), and later measures (regression path), as indicated by planned paired t tests (two-tailed), ts > 2; ps $< .05$.[2] No significant reading time differences were observed for irregular polysemes versus their controls (see also Cohen's d).

We point out that there were no reliable differences in first pass skipping rates between homonym and polyseme conditions for either the disambiguating word or the spillover region, and no interaction between Ambiguity and Word Type, zs < 2, ps $> .1$. For the Word Type factor, there was a trade off in skipping rate between regions, as disambiguating words were generally skipped more often in the ambiguous than the control conditions, $z = 4.21$, $p < .001$, whereas the opposite was true for the disambiguating word spillover region, $z = 4.78$, $p < .001$. However, this tradeoff cannot explain the sustained difference found between homonyms and polysemes.

Taken together, results for the Context After conditions are compatible with a shared representation account of irregular polysemes, where overlapping information of senses allows readers to delay full sense commitment. Reading times elicited for the disambiguating context spillover region support the view that readers accessed the dominant reading of a homonym, but not so for a polyseme. Upon encountering the ambiguous word, readers commit to a homonym's dominant meaning, but, for an irregular polyseme, they delay full sense commitment until subsequent, informative context is available. When subordinate-bias context is later encountered, readers can activate and integrate the

[2] The t test on first fixations was only marginally significant by items, $t(17) = 1.8$, $p = .089$.

Table 4
Context After Conditions: Inferential Statistics for the Context Words

Measure	Main effect/Interaction	Estimate	*SE*	*t*/*z*	*p*
First fixation	Intercept	3.90	7.66e–03	509.09	<.001
	Ambiguity	−0.16e–03	6.83e–03	−0.02	.981
	Word type	−1.73e–03	6.51e–03	−0.27	.790
	Ambiguity × Word type	7.08e–03	12.59e–03	0.56	.574
Single fixation	Intercept	3.96	8.49	465.45	<.001
	Ambiguity	−0.53	8.38	−0.06	.950
	Word type	−5.13	6.61	−0.78	.437
	Ambiguity × Word type	7.26	13.22	0.55	.583
First pass	Intercept	2.80	4.59e–03	610.02	<.001
	Ambiguity	−1.70e–03	6.22e–03	−0.27	.785
	Word type	−1.27e–03	2.93e–03	−0.43	.666
	Ambiguity × Word type	2.81e–03	5.85e–03	0.48	.631
First pass regression	Intercept	−1.80	0.14	−13.10	<.001
	Ambiguity	−0.28	0.17	−1.67	.095
	Word type	0.12	0.11	1.12	.264
	Ambiguity × Word type	0.02	0.21	0.10	.922
Regression path	Intercept	2.05	2.13e–03	958.12	<.001
	Ambiguity	−3.17e–03	2.75e–03	−1.15	.249
	Word type	1.46e–03	1.60e–03	0.91	.362
	Ambiguity × Word type	−0.21e–03	3.05e–03	−0.07	.947
Total time	Intercept	2.55	4.14e–03	617.12	<.001
	Ambiguity	−2.38e–03	5.36e–03	−0.44	.657
	Word type	3.08e–03	2.64e–03	1.17	.243
	Ambiguity × Word type	2.10e–03	5.25e–03	0.40	.689

Note. *Ambiguity* = homonym vs. polyseme; *Word type* = ambiguous vs. control word; Reading times were log-transformed using Box-Cox power transformations prior to statistical analysis (see text). Because items were not matched across the factor of ambiguity, main effects of ambiguity are not informative and are only reported for completeness.

subordinate-related unshared information for polysemes without difficulty, since the unshared sense component is compatible with the context.

Context Before Conditions

Dependent measures and Cohen's *d*s are shown in Table 6, and results of the regression models can be found in Tables 7 (Ambiguous word) and 8 (Ambiguous word spillover). Beginning with early measures, Table 7 indicates a marginal main effect of word type for first pass reading time on the ambiguous word, followed by Table 8 showing a significant main effect of word type on the spillover region for first fixations, single fixations, and first pass reading times. The significant main effect of word type in the spillover region indicates early difficulty for polysemes (like the homonyms), as predicted by the shared features model, but not the underspecification model. Table 6 shows that this early main effect is attributable to both homonyms and polysemes being more difficult than their controls. However, compared with their control words, the slow-down was beginning to be numerically stronger in single fixation and first pass durations for homonyms than polysemes. This progression is consistent with the time course of sense commitment proposed by the shared features model.

In later measures, the main effect of word type was either absent or qualified by an interaction, indicating that homonyms were processed differently than their controls, whereas polysemes were not (see also Cohen's *d* in Table 6). On the ambiguous versus control word itself (see Table 7), the reading slow-down continued for homonyms but not polysemes, shown in regression path and total time. On the spillover region (see Table 8), this difference strengthened for first pass regressions and regression path, confirmed by planned *t* tests (paired, two-tailed), ts > 3, ps $< .01$, although less so in total time. Polysemes, however, showed little to no slow-down for these same measures and regions. The overall pattern, then, is that readers took longer to integrate homonyms than polysemes, and suggests that although between-sense competition occurred for polysemes, this competition was of much smaller size for polysemes than for homonyms, also supporting our shared features model.

Importantly, longer first fixations and single fixations for polysemes in the spillover region were not compensatory. There were no reliable differences in skipping rates between homonyms (7.0%) and polysemes (9.2%) on the ambiguous word itself, $z = -0.86$, $p = .388$. Also, ambiguous words (8.1%) were overall not skipped more often than their control words (homonym controls: 6.5%; polyseme controls: 9.4%), $z = -0.08$, $p = .932$. For the ambiguous word spillover regions, we did not observe any reliable difference in skipping rates, zs < 1, ps $> .4$.[3]

Discussion

We investigated the representation and processing of irregular polysemes, comparing them to homonyms with similar frequency

[3] Note that post hoc testing also revealed no reliable effects of target word predictability (ambiguous and control) on reading time measures. Including predictability scores in the statistical models did not change the reported levels of significance (Tables 7 and 8).

Table 5
Context After Conditions: Inferential Statistics for the Context Word Spillover

Measure	Main effect/Interaction	Estimate	*SE*	*t/z*	*p*
First fixation	Intercept	4.81	14.41e–03	333.64	<.001
	Ambiguity	−22.09e–03	16.82e–03	−1.31	.189
	Word type	13.70e–03	10.89e–03	1.26	.208
	Ambiguity × Word type	33.44e–03	21.73e–03	1.54	.124
Single fixation	Intercept	4.93	15.87e–03	310.64	<.001
	Ambiguity	−28.70e–03	18.81e–03	−1.53	.127
	Word type	13.34e–03	12.83e–03	1.04	.299
	Ambiguity × Word type	31.63e–03	27.15e–03	1.17	.244
First pass	Intercept	3.12	7.08e–03	439.73	<.001
	Ambiguity	8.20e–03	9.80e–03	−0.84	.403
	Word type	9.06e–03	5.18e–03	1.75	.080
	Ambiguity × Word type	19.35e–03	9.84e–03	1.97	.049
First pass regression	Intercept	−2.26	0.13	−17.02	<.001
	Ambiguity	0.05	0.24	0.22	.825
	Word type	0.02	0.14	0.13	.895
	Ambiguity × Word type	0.15	0.28	0.52	.606
Regression path	Intercept	2.33	3.49e–03	667.66	<.001
	Ambiguity	0.96e–03	4.93e–03	−0.20	.845
	Word type	3.97e–03	2.59e–03	1.53	.126
	Ambiguity × Word type	10.27e–03	5.03e–03	2.04	.041
Total time	Intercept	3.12	8.24e–03	378.47	<.001
	Ambiguity	−6.01e–03	12.12e–03	−0.50	.620
	Word type	4.88e–03	5.19e–03	0.94	.347
	Ambiguity × Word type	11.06e–03	10.91e–03	1.01	.311

Note. *Ambiguity* = homonym vs. polyseme; *Word type* = ambiguous vs. neutral control word. Reading times were log-transformed using Box-Cox power transformations before statistical analysis (see text). Because items have not been matched across the factor of ambiguity, main effects of ambiguity are not informative and are only reported for completeness.

bias for two senses/meanings. First, we found evidence supporting a shared representation for irregular polysemes, rather than separate representations for the two senses. When subordinate-bias context appeared after the ambiguous word, homonyms showed a dominance effect compared to their controls, whereas polysemes did not. This pattern is similar to previous experiments examining regular polysemy (Frazier & Rayner, 1990; Frisson, 2009, 2015; Frisson & Pickering, 1999; Pickering & Frisson, 2001), and is in contrast to the main conclusion of Foraker and Murphy (2012).

Next, addressing our second question, we tested a shared features model of a shared sense representation. This model is posited to include activation of unshared features that carry frequency information, unlike the underspecification or core meaning accounts. Results from the Context Before conditions supported the shared features model, showing that readers experienced a reading slow-down for polysemes when preceding context and sense frequency bias supported different senses, creating between-sense competition. This point is consistent with irregular polysemes having unshared components of their senses activated.

We propose that between-sense competition arose because readers needed to make a choice as to which unshared features to activate in addition to the shared features. Preceding context sup-

Table 6
Context Before Conditions: Dependent Measures and Effect Sizes

Measure	Ambiguity	Ambiguous words: Ambiguous	Ambiguous words: Control	Ambiguous words: *d*	Ambiguous words spillover: Ambiguous	Ambiguous words spillover: Control	Ambiguous words spillover: *d*
First fixation	Homonym	252 (4)	245 (3)	.13	271 (5)	253 (4)	.27
	Polyseme	249 (3)	247 (4)	.08	261 (5)	251 (4)	.22
Single fixation	Homonym	256 (4)	246 (4)	.12	280 (6)	260 (5)	.30
	Polyseme	251 (4)	248 (4)	.08	276 (6)	261 (6)	.23
First pass	Homonym	281 (5)	266 (4)	.25	374 (10)	343 (8)	.28
	Polyseme	273 (5)	272 (5)	.09	372 (9)	363 (9)	.08
First pass regression	Homonym	18.1 (4.3)	18.7 (4.4)	.02	37.9 (5.4)	27.7 (5.0)	.44
	Polyseme	13.3 (3.8)	16.7 (4.2)	.20	33.8 (5.3)	34.9 (5.3)	.03
Regression path	Homonym	343 (7)	332 (8)	.21	568 (16)	482 (15)	.51
	Polyseme	318 (7)	339 (9)	.06	544 (16)	529 (16)	.08
Total time	Homonym	422 (10)	377 (9)	.36	530 (13)	478 (11)	.32
	Polyseme	374 (8)	377 (10)	.11	493 (12)	490 (12)	.04

Note. Standard error appears in parentheses following the mean, both in ms. First pass regressions are presented as proportions. d = Cohen's d.

Table 7
Context Before Conditions: Inferential Statistics for the Ambiguous Words

Measure	Main effect/Interaction	Estimate	*SE*	*t*/*z*	*p*
First fixation	Intercept	3.40	6.09e–03	557.63	<.001
	Ambiguity	−0.15e–03	6.46e–03	−0.02	.982
	Word type	7.02e–03	5.70e–03	1.23	.218
	Ambiguity × Word type	6.07e–03	10.94e–03	0.56	.579
Single fixation	Intercept	3.28	5.83	562.22	<.001
	Ambiguity	0.23e–03	6.26e–03	0.04	.970
	Word type	6.36e–03	5.47e–03	1.16	.245
	Ambiguity × Word type	5.59e–03	10.98e–03	0.51	.611
First pass	Intercept	2.78	3.99e–03	697.77	<.001
	Ambiguity	0.62e–03	4.73e–03	0.13	.896
	Word type	6.96e–03	4.04e–03	1.72	.085
	Ambiguity × Word type	7.31e–03	8.02e–03	0.91	.362
First pass regression	Intercept	−1.82	0.11	−16.17	<.001
	Ambiguity	0.22	0.16	1.35	.177
	Word type	−0.17	0.11	−1.55	.121
	Ambiguity × Word type	0.23	0.22	1.06	.291
Regression path	Intercept	2.08	1.98e–03	1051.85	<.001
	Ambiguity	1.92e–03	2.26e–03	0.85	.397
	Word type	1.44e–03	1.39e–03	1.04	.300
	Ambiguity × Word type	5.65e–03	2.77e–03	2.04	.041
Total time	Intercept	3.10	6.78e–03	456.89	<.001
	Ambiguity	12.00e–03	8.88e–03	1.35	.177
	Word type	14.14e–03	4.45e–03	3.17	.002
	Ambiguity × Word type	17.62e–03	8.91e–03	1.98	.048

Note. *Ambiguity* = homonym vs. polyseme; *Word type* = ambiguous vs. neutral control word. Reading times were log-transformed using Box-Cox power transformations before statistical analysis (see text). Because items have not been matched across the factor of ambiguity, main effects of ambiguity are not informative and are only reported for completeness.

ported activation of the unshared features of the subordinate sense while frequency bias supported activation of the unshared features of the dominant sense, at the polyseme. Importantly, the between-sense competition was weaker and resolved much sooner for the polysemes than homonyms, thus indicating that irregular polysemes have a shared component of their senses activated, and so are not simply like homonyms.

Taken together, then, our data suggest that, upon encountering an irregular polyseme, readers strongly activate the shared features. Activated shared features allow readers to remain uncommitted as to which sense to fully compute until informative context is available (see also Frisson, 2009). Interestingly, in our experiment, sense noncommitment occurred across a clause boundary, suggesting that it is not at all short-lived. A different picture is revealed when biasing context is available to readers before irregular polyseme encounter. Under these conditions, readers can no longer remain uncommitted at the polyseme and, instead, more fully activate the sense that is supported by preceding context. Crucially, when preceding context supports the less frequent sense, readers have good evidence for commitment to either sense: Context supports filling in the unshared features toward the less frequent sense; frequency bias supports filling in the unshared features toward the more frequent sense. It is this weighing of competing or incompatible evidence that slows retrieval down.

As we have already pointed out, the shared-features model partially resembles the core representation account advocated by Klepousniotou and her colleagues, although they applied it to regular polysemy only. In particular, Klepousniotou et al. (2008) suggested that core representations are likely to involve semantic features that are shared by all senses. However, the core representation account, unlike the shared features model, says very little about the role and contribution of features that are not shared between senses. Indeed, Klepousniotou et al. (2008) proposed that a core representation is largely associated with the dominant sense, with a subordinate sense being derived from it through a semantic rule when subordinate-biased context is available. Aside from the already mentioned observation that semantic rules are very unlikely to be at play for irregular polysemes, it is unclear whether or not application of a semantic rule is supposed to lead to observable processing costs. It is therefore difficult to evaluate the compatibility of the core representation account with the present findings.

A model that comes closer to the shared features model we suggest is that of Klepousniotou and Baum (2007), who distinguish between core features and peripheral features and who suggest that core features activation leads to underspecification and additional peripheral features activation to full retrieval of a specific sense. Although this model is compatible with our distinction between shared and unshared features, Klepousniotou and Baum (2007) did not address the role of the frequency of unshared features, something critical in our study to account for the difference in processing cost between dominant and subordinate senses. In fact, Klepousniotou and Baum (2007) propose that peripheral features collaborate toward retrieval (see also MacGregor et al., 2015), whereas our data suggest that full sense activation can lead to between-sense competition.

It should be noted that the degree of sense overlap in terms of semantic features could vary quite widely. Hence the partitioning of a word's senses into shared and unshared portions could poten-

Table 8
Context Before Conditions: Inferential Statistics for the Ambiguous Words Spillover

Measure	Main effect/Interaction	Estimate	*SE*	*t*/*z*	*p*
First fixation	Intercept	3.48	6.80e–03	511.43	<.001
	Ambiguity	2.41e–03	8.28e–03	0.29	.771
	Word type	15.19e–03	5.96e–03	2.55	.011
	Ambiguity × Word type	7.10e–03	12.05e–03	0.59	.556
Single fixation	Intercept	2.81	4.65	604.68	<.001
	Ambiguity	0.04e–03	5.99e–03	0.01	.995
	Word type	8.43e–03	3.90e–03	2.16	.031
	Ambiguity × Word type	2.21e–03	7.78e–03	0.28	.776
First pass	Intercept	3.01	6.99e–03	430.46	<.001
	Ambiguity	−9.45e–03	9.98e–03	−0.95	.344
	Word type	11.37e–03	4.58e–03	2.48	.013
	Ambiguity × Word type	12.91e–03	9.16e–03	1.41	.159
First pass regression	Intercept	−0.69	0.05	−15.28	<.001
	Ambiguity	−0.08	0.09	−0.90	.367
	Word type	0.21	0.09	2.31	.021
	Ambiguity × Word type	0.52	0.18	2.86	.004
Regression path	Intercept	4.45	23.03e–03	193.01	<.001
	Ambiguity	−19.23e–03	39.10e–03	−0.49	.623
	Word type	53.92e–03	14.37e–03	3.76	<.001
	Ambiguity × Word type	77.01e–03	26.86e–03	2.87	.004
Total time	Intercept	6.65	47.71e–03	139.40	<.001
	Ambiguity	34.14e–03	74.58e–03	0.46	.647
	Word type	74.18e–03	30.42e–03	2.44	.015
	Ambiguity × Word type	95.20e–03	60.30e–03	1.58	.114

Note. *Ambiguity* = homonym vs. polyseme; *Word type* = ambiguous vs. neutral control word. Reading times were log-transformed using Box-Cox power transformations prior to statistical analysis (see text). Because items have not been matched across the factor of ambiguity, main effects of ambiguity are not informative and are only reported for completeness.

tially range from very little shared and quite a lot unshared, to quite a lot shared and very little unshared. The shared features model we propose here predicts that the number of shared and unshared semantic features involved in retrieval is important, with more features overlapping leading to easier processing as a result of less between-sense competition. Likewise, irregular polysemes with quite a lot of the senses' features unshared are predicted to be processed and represented more like homonyms, whereas irregular polysemes with small unshared portions would be processed more like regular polysemes, experiencing little competition between senses (see also Eddington & Tokowicz, 2015; Klepousniotou et al., 2008). Note, though, that there could be two different representational ways to achieve such similar processing profiles: regular polysemes could have a very small feature overlap basis, but a very reliable, predictive, systematic relation, whereas irregular polysemes could have a larger feature overlap core, but a more idiosyncratic and less codified or predictable relation between senses.

A second prediction that falls out of a shared features model is that relative sense frequency should also play a role. Results of studies on balanced homonyms suggest that, unlike for biased homonyms, a balanced homonym's two (about) equally frequent meanings strongly compete with each other during lexical access. For example, when there is no preceding constraining context, gaze duration times (or other early processing measures) for balanced homonyms are longer than for matched unambiguous control words (Duffy et al., 1988; Folk & Morris, 2003; Rayner & Duffy, 1986; see also Rayner & Frazier, 1989). This finding suggests that the frequency of use of a homonym's meanings affects the degree to which these meanings compete during retrieval.

Relatedly, an interesting question for future research is whether larger versus smaller dominance scores (e.g., .90 vs. .65) would affect the strength of dominance effects for irregular polysemes. Faster reading times for larger dominance has been demonstrated for homonyms (e.g., Duffy et al., 1988), and is suggested by Foraker and Murphy's (2012) findings that a higher degree of dominance tended to produce stronger dominance effects in self-paced and eye-tracking measures of reading. To our knowledge, no experiments have examined balanced polysemes explicitly. Foraker and Murphy (2012) included a wide range of sense frequency from biased through balanced senses, but did not break down analysis in this manner. Similarly, Klepousniotou (2002; Klepousniotou et al., 2008) included a range of sense frequency in her investigations of sense priming during lexical decisions, but did not examine that variable specifically. We expect that at retrieval, like for balanced homonyms, senses of balanced irregular polysemes should compete for activation. Crucially, between-sense competition should vary as a function of unshared features activation. Little competition should incur when there is large sense overlap and strong competition should incur when there is weak sense overlap. In other words, just like increasing unshared features activation is predicted to lead to increasing dominance effects for biased irregular polysemes, increasing unshared features activation is predicted to lead to increasing between-sense competition for balanced irregular polysemes. A strong test of this hypothesis would include biased and balanced irregular polysemes with various degrees of semantic overlap.

Our data fit nicely with the observation that readers do not always compute full representations during reading, but rather retrieve "good-enough" representations (Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007). For biased homonyms, good-enough representations fall within dominant meanings, as these meanings are most frequent. For irregular polysemes (without preceding biasing context), on the other hand, good-enough representations fall within shared representations, which are proposed to be most frequent (as they are activated regardless of the intended sense).

Finally, we should reiterate that the underspecification model has, so far, not been tested experimentally on irregular polysemes. Indeed, the underspecification model has explicitly been developed to account for reading times of regular polysemes, such as *book* and *Vietnam* (Frisson, 2015; Frisson & Frazier, 2005; Frisson & Pickering, 1999; Pickering & Frisson, 2001; see Frisson, 2009 for a review). But, reading time data for regular object-content polysemes (e.g., *book*, *poem*, and *journal*) recently presented in Frisson (2015) are seemingly compatible with the shared features model proposed here. Like in our experiment, Frisson found no reliable reading time differences for dominant- versus subordinate-bias context words when they followed ambiguous words. Also, when context words preceded the ambiguous words, readers fixated marginally longer on polysemes when they followed subordinate-bias contexts than when they followed dominant-bias contexts. For a very similar condition, Frazier and Rayner (1990) also found that reading times on polysemes were marginally (polyseme) or significantly (polyseme + spillover) longer than the controls for one early measure (first pass, corrected for letter length), but the slow-down did not occur in later measures (total time, also regressions out).

Prima facie, the suggestive dominance effects found by Frisson (2015) and Frazier and Rayner (1990), which are comparable in time course to our data, appear to support a shared features type of representation rather than an underspecification account. However, we suggest that dominance effects of sense frequency are driven by the unshared portion of a sense, not the shared/core/common component. A major question still unresolved, then, is what an underspecified representation *does* consist of, which we leave for further investigation (see Frisson, 2009 for additional discussion).

Even when a continuum of relatedness is based on objective ratings (Klepousniotou et al., 2008), the assumption that relatedness and semantic feature overlap are the same thing may not be warranted, particularly for regular polysemy. Murphy and colleagues have highlighted the distinction between semantic *similarity* of senses, which is the essence of a core meaning, but not necessarily the same as *relatedness* of senses (Foraker & Murphy, 2012; Klein & Murphy, 2001, 2002; Murphy, 2007). For example, the "squawking chicken" and "roasted chicken" senses are not as conceptually similar as one might first assume. Despite the fact that one arises from the other in a reliable and systematic fashion—animal-meat relatedness—the degree of semantic overlap in terms of semantic features that make up the concepts is quite limited—one is animate, the other not; they look different, smell different, feel different, sound different, have different functions, and so on. Therefore, regular polysemy may indeed produce high relatedness ratings, but the basis of relation is very limited, and there may not actually be many semantic features overlapping or shared in common. Hence, although there is an objective continuum of relatedness, we note that the continuum is not necessarily based on the same mechanism—high (perceived) relatedness does not necessarily entail high semantic similarity or feature overlap, whereas moderate relatedness may well be attributable to semantic feature overlap.

A strong test of the underspecification and shared features models with respect to the core meaning assumption and its criticism, as well as the regular polysemy versus irregular polysemy distinction, would be to include a sufficiently controlled and judiciously selected set of regular and irregular polysemes in a within-subject design. A promising experimental approach for detecting theory-driven differences quantitatively has been presented in the current paper.

References

Apresjan, J. (1974). Regular polysemy. *Linguistics, 142,* 5–32.

Armstrong, B. C., & Plaut, D. C. (2008). Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence. *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 273–278). Hillsdale, NJ: Cognitive Science Society.

Armstrong, B. C., & Plaut, D. C. (2011). Inducing homonymy effects via stimulus quality and (not) nonword difficulty: Implications for models of semantic ambiguity and word recognition. *Proceedings of the 33rd annual conference of the Cognitive Science Society* (pp. 2223–2228). Hillsdale, NJ: Cognitive Science Society.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–6. http://CRAN.R-project.org/package=lme4

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: An MEG study. *Cognitive Brain Research, 24,* 57–65. http://dx.doi.org/10.1016/j.cogbrainres.2004.12.006

Binder, K. S., & Morris, R. K. (1995). Eye movements and lexical ambiguity resolution: Effects of prior encounter and discourse topic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1186–1196. http://dx.doi.org/10.1037/0278-7393.21.5.1186

Binder, K. S., & Morris, R. K. (2011). An eye-movements analysis of ambiguity resolution: Beyond meaning access. *Discourse Processes, 48,* 305–330. http://dx.doi.org/10.1080/0163853X.2011.577391

Binder, K. S., & Rayner, K. (1998). Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations and selfpaced reading. *Psychonomic Bulletin & Review, 5,* 271–276. http://dx.doi.org/10.3758/BF03212950

Binder, K. S., & Rayner, K. (1999). Does contextual strength modulate the subordinate bias effect? A reply to Kellas and Vu. *Psychonomic Bulletin & Review, 6,* 518–522. http://dx.doi.org/10.3758/BF03210843

Blanchard, H. E., Pollatsek, A., & Rayner, K. (1989). The acquisition of parafoveal word information in reading. *Perception & Psychophysics, 46,* 85–94. http://dx.doi.org/10.3758/BF03208078

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, Instruments & Computers, 41,* 977–990. http://dx.doi.org/10.3758/BRM.41.4.977

Clark, H. H., & Gerrig, R. J. (1983). Understanding old words with new meanings. *Journal of Verbal Learning & Verbal Behavior, 22,* 591–608. http://dx.doi.org/10.1016/S0022-5371(83)90364-X

Colbert-Getz, J., & Cook, A. E. (2013). Revisiting effects of contextual strength on the subordinate bias effect: Evidence from eye movements. *Memory & Cognition, 41,* 1172–1184. http://dx.doi.org/10.3758/s13421-013-0328-3

Copestake, A., & Briscoe, E. J. (1995). Semi-productive polysemy and sense extension. *Journal of Semantics, 12,* 15–67. http://dx.doi.org/10.1093/jos/12.1.15

Cruse, D. A. (1986). *Lexical semantics*. Cambridge, UK: Cambridge University Press.

Dopkins, S., Morris, R. K., & Rayner, K. (1992). Lexical ambiguity and eye fixations in reading: A test of competing models of lexical ambiguity resolution. *Journal of Memory and Language, 31,* 461–476. http://dx.doi.org/10.1016/0749-596X(92)90023-Q

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27,* 429–446. http://dx.doi.org/10.1016/0749-596X(88)90066-6

Eddington, C. M., & Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: The current state of the literature. *Psychonomic Bulletin & Review, 22,* 13–37. http://dx.doi.org/10.3758/s13423-014-0665-7

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11,* 11–15. http://dx.doi.org/10.1111/1467-8721.00158

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass, 1,* 71–83. http://dx.doi.org/10.1111/j.1749-818X.2007.00007.x

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition, 96,* 263–284. http://dx.doi.org/10.1016/j.cognition.2004.09.002

Fishbein, J., & Harris, J. A. (2014). Making sense of Kafka: Structural biases induce early sense commitment for metonyms. *Journal of Memory and Language, 76,* 94–112. http://dx.doi.org/10.1016/j.jml.2014.06.005

Folk, J. R., & Morris, R. K. (2003). Effects of syntactic category assignment on lexical ambiguity resolution in reading: An eye movement analysis. *Memory & Cognition, 31,* 87–99. http://dx.doi.org/10.3758/BF03196085

Foraker, S., & Murphy, G. L. (2012). Polysemy in sentence comprehension: Effects of meaning dominance. *Journal of Memory and Language, 67,* 407–425. http://dx.doi.org/10.1016/j.jml.2012.07.010

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language, 29,* 181–200. http://dx.doi.org/10.1016/0749-596X(90)90071-7

Frisson, S. (2009). Semantic underspecification in language processing. *Language and Linguistics Compass, 3,* 111–127. http://dx.doi.org/10.1111/j.1749-818X.2008.00104.x

Frisson, S. (2015). About bound and scary books: The processing of book polysemies. *Lingua, 157,* 17–35. http://dx.doi.org/10.1016/j.lingua.2014.07.017

Frisson, S., & Frazier, L. (2005). Carving up word meaning: Portioning and grinding. *Journal of Memory and Language, 53,* 277–291. http://dx.doi.org/10.1016/j.jml.2005.03.004

Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1366–1383. http://dx.doi.org/10.1037/0278-7393.25.6.1366

Frisson, S., & Pickering, M. J. (2001). Obtaining a figurative interpretation of a word: Support for underspecification. *Metaphor and Symbol, 16,* 149–171. http://dx.doi.org/10.1080/10926488.2001.9678893

Frisson, S., & Pickering, M. J. (2007). The processing of familiar and novel senses of a word: Why reading Dickens is easy but reading Needham can be hard. *Language and Cognitive Processes, 22,* 595–613. http://dx.doi.org/10.1080/01690960601017013

Frisson, S., Pickering, M. J., & McElree, B. (2011). The difficult mountain: Enriched composition in adjective-noun phrases. *Psychonomic Bulletin & Review, 18,* 1172–1179. http://dx.doi.org/10.3758/s13423-011-0142-5

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

Gentner, D., Bowdle, B. F., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 199–253). Cambridge, MA: MIT Press.

Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics, 31,* 919–929. http://dx.doi.org/10.1016/S0378-2166(98)00100-3

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences, 7,* 92–96. http://dx.doi.org/10.1016/S1364-6613(02)00040-2

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review, 97,* 3–18. http://dx.doi.org/10.1037/0033-295X.97.1.3

Hino, Y., Pexman, P. M., & Lupker, S. J. (2006). Ambiguity and relatedness effects in semantic tasks: Are they due to semantic coding? *Journal of Memory and Language, 55,* 247–273. http://dx.doi.org/10.1016/j.jml.2006.04.001

Kambe, G., Rayner, K., & Duffy, S. A. (2001). Global context effects on processing lexically ambiguous words: Evidence from eye fixations. *Memory & Cognition, 29,* 363–372. http://dx.doi.org/10.3758/BF03194931

Kellas, G., & Vu, H. (1999). Strength of context *does* modulate the subordinate bias effect: A reply to Binder and Rayner. *Psychonomic Bulletin & Review, 6,* 511–517. http://dx.doi.org/10.3758/BF03210842

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review, 7,* 257–266. http://dx.doi.org/10.3758/BF03212981

Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language, 45,* 259–282. http://dx.doi.org/10.1006/jmla.2001.2779

Klein, D. E., & Murphy, G. L. (2002). Paper has been my ruin: Conceptual relations of polysemous senses. *Journal of Memory and Language, 47,* 548–570. http://dx.doi.org/10.1016/S0749-596X(02)00020-7

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language, 81,* 205–223. http://dx.doi.org/10.1006/brln.2001.2518

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics, 20,* 1–24. http://dx.doi.org/10.1016/j.jneuroling.2006.02.001

Klepousniotou, E., Pike, G. B., Steinhauer, K., & Gracco, V. (2012). Not all ambiguous words are created equal: An EEG investigation of homonymy and polysemy. *Brain and Language, 123,* 11–21. http://dx.doi.org/10.1016/j.bandl.2012.06.007

Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1534–1543. http://dx.doi.org/10.1037/a0013012

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago, IL: University of Chicago Press. http://dx.doi.org/10.7208/chicago/9780226471013.001.0001

Langacker, R. W. (1987). *Foundations of cognitive grammar*. Stanford, CA: Stanford University Press.

Lehrer, A. (1990). Polysemy, conventionality, and the structure of the lexicon. *Cognitive Linguistics, 1,* 207–246. http://dx.doi.org/10.1515/cogl.1990.1.2.207

MacGregor, L. J., Bouwsema, J., & Klepousniotou, E. (2015). Sustained meaning activation for polysemous but not homonymous words: Evidence from EEG. *Neuropsychologia, 68,* 126–138. http://dx.doi.org/10.1016/j.neuropsychologia.2015.01.008

Martin, C., Vu, H., Kellas, G., & Metcalf, K. (1999). Strength of discourse context as a determinant of the subordinate bias effect. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 52,* 813–839. http://dx.doi.org/10.1080/713755861

Murphy, G. L. (2007). Parsimony and the psychological representation of polysemous words. In M. Rakova, G. Pethö, & C. Rákosi (Eds.), *The cognitive basis of polysemy* (pp. 47–70). Frankfurt am Maim, Germany: Peter Lang Verlag.

Navarro, D. J. (2014). *Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.4)*. Adelaide, Australia: University of Adelaide.

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy, 3,* 143–184. http://dx.doi.org/10.1007/BF00126509

Nunberg, G. (1995). Transfers of meaning. *Journal of Semantics, 12,* 109–132. http://dx.doi.org/10.1093/jos/12.2.109

Pickering, M. J., & Frisson, S. (2001). Processing ambiguous verbs: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 556–573. http://dx.doi.org/10.1037/0278-7393.27.2.556

Pickering, M. J., Frisson, S., McElree, B., & Traxler, M. J. (2004). Eye movements and semantic composition. In M. Carreiras, Jr., & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs, and beyond. Hove,* UK: Psychology Press.

Pickering, M. J., McElree, B., Frisson, S., Chen, L., & Traxler, M. (2006). Underspecification and aspectual coercion. *Discourse Processes, 42,* 131–155. http://dx.doi.org/10.1207/s15326950dp4202_3

Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.

Pylkkänen, L., Llinás, R., & Murphy, G. L. (2006). The representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience, 18,* 97–109. http://dx.doi.org/10.1162/089892906775250003

Pylkkänen, L., Martin, A. E., McElree, B., & Smart, A. (2009). The anterior midline field: Coercion or decision making? *Brain and Language, 108,* 184–190. http://dx.doi.org/10.1016/j.bandl.2008.06.006

Pylkkänen, L., & McElree, B. (2006). The syntax-semantics interface: On-line composition of sentence meaning. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 539–577). New York, NY: Elsevier. http://dx.doi.org/10.1016/B978-012369374-7/50015-8

Rabagliati, H., Marcus, G. F., & Pylkkänen, L. (2010). Shifting senses in lexical semantic development. *Cognition, 117,* 17–37. http://dx.doi.org/10.1016/j.cognition.2010.06.007

Rabagliati, H., Marcus, G. F., & Pylkkänen, L. (2011). Rules, radical pragmatics and restrictions on regular polysemy. *Journal of Semantics, 28,* 485–512. http://dx.doi.org/10.1093/jos/ffr005

Rabagliati, H., & Snedeker, J. (2013). The truth about chickens and bats: Ambiguity avoidance distinguishes types of polysemy. *Psychological Science, 24,* 1354–1360. http://dx.doi.org/10.1177/0956797612472205

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. http://dx.doi.org/10.1037/0033-2909.124.3.372

Rayner, K., Binder, K. S., & Duffy, S. A. (1999). Contextual strength and the subordinate bias effect: Comment on Martin, Vu, Kellas, and Metcalf. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 52,* 841–852. http://dx.doi.org/10.1080/713755868

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14,* 191–201. http://dx.doi.org/10.3758/BF03197692

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 779–790. http://dx.doi.org/10.1037/0278-7393.15.5.779

Rayner, K., Pacht, J. M., & Duffy, S. A. (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language, 33,* 527–544. http://dx.doi.org/10.1006/jmla.1994.1025

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rice, S. A. (1992). Polysemy and lexical representation: The case of three English prepositions. In *Proceedings of the 14th conference of the Cognitive Science Society* (pp. 89–94). Hillsdale, NJ: Erlbaum.

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

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2004). Modeling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28,* 89–104. http://dx.doi.org/10.1207/s15516709cog2801_4

Ruhl, C. (1989). *On monosemy: A study in linguistic semantics*. Albany, NY: SUNY Press.

Sereno, S. C., O'Donnell, P. J., & Rayner, K. (2006). Eye movements and lexical ambiguity resolution: Investigating the subordinate-bias effect. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 335–350.

Simpson, G. B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning & Verbal Behavior, 20,* 120–136. http://dx.doi.org/10.1016/S0022-5371(81)90356-X

Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 28–39. http://dx.doi.org/10.1037/0096-1523.11.1.28

Simpson, G. B., & Krueger, M. A. (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language, 30,* 627–643. http://dx.doi.org/10.1016/0749-596X(91)90029-J

Srinivasan, M., & Snedeker, J. (2011). Judging a *book* by its cover and its contents: The representation of polysemous and homophonous meanings in four-year-old children. *Cognitive Psychology, 62,* 245–272. http://dx.doi.org/10.1016/j.cogpsych.2011.03.002

Taylor, J. R. (2003). Polysemy's paradoxes. *Language Sciences, 25,* 637–655. http://dx.doi.org/10.1016/S0388-0001(03)00031-7

Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language, 47,* 530–547. http://dx.doi.org/10.1016/S0749-596X(02)00021-9

Tuggy, D. (1993). Ambiguity, polysemy, and vagueness. *Cognitive Linguistics, 4,* 273–290. http://dx.doi.org/10.1515/cogl.1993.4.3.273

Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory & Cognition, 22,* 111–126. http://dx.doi.org/10.3758/BF03202766

Vu, H., Kellas, G., & Paul, S. T. (1998). Sources of sentence constraint on lexical ambiguity resolution. *Memory & Cognition, 26,* 979–1001. http://dx.doi.org/10.3758/BF03201178

Williams, J. N. (1992). Processing polysemous words in context: Evidence for interrelated meanings. *Journal of Psycholinguistic Research, 21,* 193–218. http://dx.doi.org/10.1007/BF01068072

Received April 14, 2015
Revision received February 1, 2016
Accepted February 3, 2016 ■