Semantic Ambiguity Effects in Lexical Processing: A Neural-Network Account Based on Semantic Settling Dynamics

Blair C. Armstrong  
Basque Center on Cognition, Brain, and Language  
David C. Plaut  
Carnegie Mellon University

The meanings of most words depend on the context in which they occur (e.g., <river> vs. <money> BANK). Developing a theory of how comprehension of semantically ambiguous words are understood is a critical aspect of any theory of word or discourse comprehension. However, success to date has been limited by discrepancies in the effects of relatedness of meaning observed within and between tasks. Further, existing accounts are underspecified, narrow in scope, and mutually inconsistent. The current work introduces the semantic settling dynamics account of semantic ambiguity resolution, in which the discrepant effects are explained by the temporal settling dynamics in semantics within a neural network, and how these dynamics interact with the semantic representations of ambiguous words over time. This account stands as an alternative to one based on the configuration of the decision system across tasks (Hino, Pexman, & Lupker, 2006, JML). The proposed account reconciles a wide body of disparate results within a single unified mechanistic account, is supported by initial investigations that vary processing time to modulate semantic ambiguity effects, and generates targeted predictions for future computational, neural, and behavioral research.

Keywords: semantic ambiguity; task differences; temporal processing dynamics; semantic settling dynamics; connectionist models

1. Introduction

Developing a comprehensive mechanistic account of word comprehension—that is, how a word’s orthographic and phonological representations activate a semantic representation—is a central goal for researchers studying word and discourse processing. Given that many words are associated with multiple distinct interpretations depending on the context in which they occur, one of the core challenges that must be addressed is how context-sensitive word comprehension occurs. For instance, BANK may refer to a financial institution or to the edge of a body of water,¹ and PAPER may refer to an academic work or the material loaded into a printer (these semantic ambiguities are denoted later in the form <river>/<money> BANK). In fact, the vast majority of the content words in English and other languages are semantically ambiguous—one estimate indicates that they comprise over 80% of high frequency content words (Klein & Murphy, 2001; see also Beretta, Fiorentino, & Poeppel, 2005). This is illustrated in Figure 1, which presents the distribution of words as a function of the number of unrelated meanings (e.g., the two interpretations of BANK) and of the number of related senses (e.g., the two interpretations of PAPER, which may share some common features), using definitions in the Wordsmyth Online Dictionary—a source that has been shown previously to provide meaning and sense estimates that correlate with psychological measures (Azuma & Van Orden, 1997; Rodd, Gaskell, & Marslen-Wilson, 2002). Based on these estimates, it is clear that developing a theory of how semantically ambiguous words are processed is critical to a full understanding of word comprehension, and to bridging between isolated word and discourse comprehension (Cairns, Waltzman, & Schlisselberg, 2004; McClelland, St John, & Taraban, 1989).

A large body of research has been devoted to the study

¹Such words are sometimes called “lexically” ambiguous in the case of homonyms with unrelated meanings, but this terminology presumes that each meaning is denoted by a separate entry in a mental lexicon—a theoretical distinction that may not hold. Given that such items are also “semantically” ambiguous, we will use the latter term throughout.
Figure 1. Scatterplot of the number of unrelated meanings and number of related senses associated with 47,404 entries in the Wordsmyth Online Dictionary (Parks, Ray, & Bland, 1998). For consistency with later sections, words with a single meaning and a single sense are labeled as unambiguous, words with only one related sense associated with each unrelated meaning are labeled as homonymous, words with multiple related senses associated with a single meaning are labeled as polysemous, and words with multiple unrelated meanings, one or more of which is associated with multiple related senses, are labeled as hybrid ambiguous. Individual data points have been jittered within each cell to better illustrate the total number of items.

We aim to advance the semantic ambiguity literature on several fronts. First, we present a critical review of the core findings in the literature to organize the vast literature on the subject and to establish the key phenomena that a general theory of ambiguous word comprehension should explain. Second, the strengths and weaknesses of prior accounts of semantic ambiguity effects are considered in relation to this literature, with the aim of identifying the role that these theories, or updated variants thereof, may play in developing a more comprehensive account. Next, we introduce a novel account of these phenomena, the semantic settling dynamics (SSD) account, which is based on representational differences and temporal processing dynamics within a biologically-motivated connectionist network. This account is substantiated via additional analyses of prior simulations that were used to provide an initial assessment of the account, which show that this account captures an unprecedentedly wide set of phenomena, as identified in the first portion of the review. The account is also used to motivate targeted empirical studies and re-analyses of some past behavioral work, which provide further evidence consistent with the proposal and which are not captured by other prior accounts. Finally, the implications of the work are discussed in relation to the semantic ambiguity literature, discourse comprehension, and the value of computationally-explicit, biologically-motivated models.

2. Review of the Semantic Ambiguity Literature

The review aims to identify the typical patterns of results that have been observed across a range of tasks. Where relevant, we discuss likely causes for some of the discrepancies in what has been reported, particularly in the context of highly-similar tasks. We first consider studies of the processing of ambiguous words in isolation before discussing how such items are processed in biasing contexts. In addition, for the sake of clarity, the relatedness-of-interpretations continuum that characterizes the relatedness among an ambiguous word’s different interpretations was subdivided into two main groups: ambiguous words with relatively unrelated interpretations, termed homonyms (e.g., BANK), and ambiguous words with related interpretations, termed polysemes (e.g., PAPER). Words that have both related and unrelated interpretations are referred to as hybrid words. These categories of ambiguous words are often compared both against one another and against unambiguous control words that typically have only a single interpretation, or at least a small number of interpretations in comparison to homonyms and polysemes.

2.1 Ambiguous words in isolation

One of the most frequently employed tasks for investigating semantic ambiguity, and lexical processing in general, is the lexical decision task (e.g., Borowsky & Masson, 1996; Jastrzembski, 1981; Klepousniotou & Baum, 2007; Millis & Button, 1989; Rodd et al., 2002; Rubenstein et al., 1970). This task may be interpreted as assessing how words activate...
semantic representations when there is no systematic pressure to select a particular interpretation if the word is ambiguous, provided that the nonword foils are sufficiently wordlike to rule out surface-level discriminations (e.g., Joordens & Besner, 1994; Piercey & Joordens, 2000; Rodd, Gaskell, & Marslen-Wilson, 2004, although for alternative views, see Balota & Chumbley, 1984; Hino & Lupker, 1996; Seidenberg, Waters, Sanders, & Langer, 1984). Insofar as a lexical decision can be made based on very coarse semantic information—for instance, whether a given letter string associated with any meaning whatsoever—the lexical decision task may also be generally thought of as probing relatively early semantic access (Piercey & Joordens, 2000; Masson & Borowsky, 1995). This is concordant with the relatively rapid latencies typically reported in behavioral lexical decision studies, and with the ERP correlates of meaning access that could have been tapped at the moment a response is initiated (Laszlo & Fedrmeier, 2014) which is expected to be 100-200 ms prior to the motor act of pressing a response button.

In many earlier studies and in a small number of recent studies, the results of lexical decision experiments indicate an overall advantage for ambiguous items. However, most of these studies either suffered from confounds with other factors that influence lexical decision performance (e.g., familiarity, see Gernsbacher, 1984), or failed to distinguish homonymy and polysemy (e.g., Borowsky & Masson, 1996; Hino & Lupker, 1996; Kellas, Ferraro, & Simpson, 1988; Millis & Button, 1989; although Hino et al., 2010, argue that the earlier Hino & Lupker study does not suffer from this confound). We thus focus on studies that address these issues in their experimental designs and/or statistical methods.

2.1.1 Relatedness of meanings. In an influential study, Rodd, Gaskell, and Marslen-Wilson (2002) reanalyzed the experimental items from prior studies by Millis and Button (1989), Borowsky and Masson (1996, see also Piercey & Joordens, 2000, for a replication and extension of these results), and Azuma and Van Orden (1997). Noting a potential confound, they hypothesized that the previously reported ambiguity advantage could, in fact, be restricted to the polysemes only. Rodd and colleagues tested this hypothesis in two visual lexical decision experiments and observed a consistent processing advantage for all items with many related senses (polysemes and hybrid words) relative to homonyms (Experiment 1), and homonyms and unambiguous controls (Experiment 2; see also similar results reported by Armstrong & Plaut, 2008 and Locker, Simpson, & Yates, 2003). These effects were observed only in the context of pseudohomophone nonword foils, suggesting that these types of foils may be especially well suited for maximizing nonword wordlikeness and, consequently, semantic contributions to the decision system (see also Azuma & Van Orden, 1997, but cf. Armstrong & Plaut, 2008, 2011, for discussion of why orthographically-wordlike non-pseudohomophones may be superior foils). These studies thus provided convergent evidence that the performance difference between these word types relates to a psychological difference in terms of how these items are processed.

In a third experiment using auditory lexical decision and a subset of the items from their previous studies, Rodd et al. (2002) observed a weaker but statistically significant polysemy advantage as well as a strong homonymy disadvantage. Of particular relevance in the present context, the overall latencies in the auditory task were approximately 350 ms slower than those in the visual tasks. Although part of this latency increase is undoubtedly due to the time needed to present each auditory stimulus, recent research also suggests that the temporal dynamics of auditory (as compared to visual) word recognition may allow for additional semantic processing to take place (Armstrong, Barrio Abad, & Samuel, 2014).

The results reported by Tamminen, Cleland, Quinlan, and Gaskell (2006) using the Psychological Refractory Period (PRP) paradigm support this perspective (although the authors explain their results in terms of lexical vs. semantic processing loci). Once more, this study involved the same word stimuli and auditory lexical decision task reported by Rodd et al. (2002). However, on each trial, before the lexical stimulus relevant to the lexical decision task was presented, participants were first presented with a colored shape upon which they needed to perform a color discrimination task. Varying the stimulus onset asynchrony (SOA) for these two tasks (100 vs. 1000 ms) allowed the experimenters to manipulate the overall latencies for the lexical decision task. Overall latencies for words at the long SOA were similar to those reported in the Rodd et al. (2002) study (883 ms), whereas they were considerably longer (1281 ms) in the case of the short SOA. With respect to the ambiguity effects, the authors observed an equivalently large polysemy advantage at both SOAs, but observed a homonymy disadvantage only at the long SOA. Thus, although these results do not perfectly replicate the specific pattern of ambiguity effects in the standalone auditory lexical decision task reported by Rodd and colleagues (2002; i.e., strong evidence for a homonymy disadvantage, weak evidence for a polysemy disadvantage), they are consistent with the notion that relatively slower responses are associated with a consistent and different pattern of responses than for fast responses.

Aiming to understand the neural underpinnings of the results obtained in Rodd et al.’s (2002) second behavioral study, Beretta et al. (2005) replicated the experiment and used magnetoencephalography (MEG) concurrently to record the neural correlates associated with performing the task in an event-related design. In the behavioral data, they observed slightly slower overall latencies and both a homonymy disadvantage and a polysemy advantage. The
MEG data also showed a significant processing advantage for words with many senses and a significant processing disadvantage for words with many meanings, as reflected by differences in peak latencies on the M350 component. There was also a trend for greater amplitude differences on the M350 for words with many meanings. The M350 component is analogous to the N350/N400 component of event-related brain potentials (ERPs), which is linked with a source in the anterior temporal lobe that subserves semantic processing (Kutas & Federmeier, 2011; McClelland & Rogers, 2003; and which may also form part of a broader semantic network that includes the inferior frontal gyrus, angular gyrus, middle temporal gyrus, and superior temporal gyrus; for a review, see Lau, Phillips, & Poeppel, 2008), suggesting a semantic basis for these effects. Additionally, this component has a long history of being relatively insensitive to later processes such as post-lexical decision making (Pykkänen, Stringfellow, & Marantz, 2002). Similarly, no effects of ambiguity were reported on earlier components typically associated with the processing of the surface characteristics of a word, such as orthography and phonology (e.g., N170, M250; for a similar lack of effects on earlier components, see Simon, Lewis, & Marantz, 2012; Taler, Klepousniotou, & Phillips, 2009).

Collectively then, Beretta et al.’s (2005) results suggest that semantic processing is driving the behavioral effects (although see Taler et al., 2009 for a decoupling of N400 and behavioral ambiguity effects in individuals with mild cognitive impairment). However, certain aspects of their methodology reduces the weight that should be placed on these results. First, their methods of filtering participants based on whether they showed a clear M350 component may have created a selection bias toward individuals who generate particularly strong semantic activity. Second, the authors failed to include in their by-item analyses covariates such as frequency and familiarity that have been shown to eliminate significant homonymy effects in replication studies with this particular item set (Armstrong & Plaut, 2011). More broadly, Armstrong and Plaut’s (2011) failure to replicate the behavioral effects reported by Rodd et al. (2002) also calls into question the robustness of the original a homonymy disadvantage, although dialect differences between those two studies (British vs. American English) may contribute to this discrepancy. Indeed, recent work comparing a large set of 578 homonyms in the European and Rionatelene dialects of Spanish showed that there is considerable variability in participants’ estimated variability of use of different interpretations across the dialects (Armstrong, Zugarramurdi, Cabana, Valle Lisboa, & Plaut, in press). In that study, the authors reported that 72% of the variance in estimates of the relative frequency of a given meaning of an ambiguous word were shared across dialects. Moreover, across the two dialects only 28% of the variance was explained in the correlation between whether participants knew any meaning of a given homonym. Collectively, these findings highlight the critical importance of dialect-specific controls in ensuring sensitive and statistically-powerful evaluations of the effects homonymy, and of the potential importance of graded variations in the relative meaning frequency of a homonym’s meanings.

2.1.2 Meaning dominance. Related work has examined the role of meaning dominance in modulating performance across the continuous range of meaning relatedness between homonymy and polysemy. Klepousniotou and Baum (2007) investigated the effects of meaning dominance (balanced vs. unbalanced meaning frequencies) on homonymy effects, and of increased sense overlap on polysemy effects. Sense overlap was established using the linguistic distinction between metaphoric polysemy (e.g., <celestial>/<movie> STAR), which was assumed to involve less overlapping and more idiosyncratically related senses, and metonymic polysemy (e.g., <animal>/<meat> RABBIT), which was assumed to involve more overlapping and regularized relationships among senses. In both auditory and visual lexical decision, they observed a processing advantage for metonymically polysemous items relative to metaphorically polysemous items, unambiguous controls, and homonyms. They also reported a processing advantage for metaphorically polysemous items in the auditory task. These behavioral results parallel those obtained in other work that showed the polysemy effect is enhanced by both the number and relatedness of a polyseme’s interpretations (Locker et al., 2003), as well as in electrophysiological measures of N400 amplitudes in healthy aged participants (Taler et al., 2009), which further suggests a semantic basis for these effects. Interestingly, however, no effects of homonymy were detected in either task, even in the context of balanced homonyms that would presumably generate the strongest competition between meanings. This agrees with the visual lexical decision results reported by Rodd et al. (2002), but disagrees with the significant homonymy disadvantage reported in visual lexical decision by Beretta et al. (2005) and in auditory lexical decision with roughly similar overall latencies by Rodd and colleagues.

2.1.3 Grammatical category. Mirman et al. (2010) proposed a potential explanation for the inconsistent reports of a homonymy disadvantage. They examined whether the effects of homonymy for balanced homonyms depended on whether the homonyms had either noun-noun or noun-verb meanings (e.g., <card>/<boat> DECK vs. <dog>/<tree> BARK). Given that noun-verb homonyms do not share grammatical class and the noun interpretations often do not have the action representations typically associated with verbs
(Watson, 2009), these items may have less semantic overlap than noun-noun homonyms. Consequently, they should generate greater competition among the inconsistent features of each interpretation and, by proxy, a stronger homonymy disadvantage. Surprisingly, this was not the case in the empirical data: noun-noun homonyms were significantly slower relative to unambiguous controls, whereas responses to the noun-verb homonyms were not significantly slower than those to the unambiguous items in the item analyses (although the analyses did not include any covariates, leaving open other possible explanations of these effects; for alternative perspectives on what variables to control for and how this might be achieved, see Armstrong & Plaut, 2011; Armstrong, Watson, & Plaut, 2012; Gernsbacher, 1984). Similar results were also obtained in a four-choice picture-word matching task in which participants were presented with a spoken word and eye movement latencies were recorded to a target image related to the meaning of the word, along with three other distractor pictures. The eye-tracking experiment additionally showed that eye movements to the target image for trials involving noun-verb homonyms were both significantly faster than the eye movements for trials involving noun-noun homonyms and significantly slower than the eye movements for trials involving unambiguous controls. Collectively, these results led Mirman and colleagues to propose that there is a non-monotonic modulation of competitive dynamics as a function of the distance between competing representations.

Setting aside the detailed theoretical claims regarding the basis for the performance differences between noun-noun and noun-verb homonyms, the results reported by Mirman et al. (2010) clearly point to grammatical class as an important covariate that could explain the discrepant homonymy effects reported in some prior studies. An inspection of the balanced homonyms used by Klepousniotou and Baum (2007) and the homonyms used by Rodd et al. (2002) indicated that most of the items in those studies had both noun and verb interpretations—of the 18 items in the balanced homonym condition in the study conducted by Klepousniotou and Baum (2007), 10 had both noun and verb interpretations in the Wordsmyth dictionary (Parks et al., 1998). All of the 23 items in the auditory lexical decision task reported by Rodd et al. (2002), which represented the majority of the items in their second visual lexical decision experiment, also had at least one noun and one verb meaning. Thus, the use of noun-verb homonyms, which generate weaker homonymy effects, is likely a contributing factor to the inconsistent effects of homonymy across different lexical decision experiments.

2.1.4 Concreteness and other psycholinguistic variables. Paralleling the investigations into how other psycholinguistic properties such as grammatical class may mediate the homonymy effects, recent work by Jager and Cleland (2014) has also probed how other semantic properties, such as concreteness, may mediate polysemy effects. They reported two studies, each involving carefully matched sets of unambiguous words and polysemes, but that differed in terms of concreteness. In the case of the very concrete items, no polysemy effects were detected; however, the less concrete items did show a polysemy advantage. Given that previous item sets (e.g., Rodd et al., 2002) have included items that span a range of concreteness values, the authors conclude that polysemy and concreteness interact, and that previous reports of an overall polysemy advantage may have been driven by the less concrete items in each subset. Via additional regression analyses, they also rule out alternative explanations of these results based strictly on semantic diversity, which measures how many different contexts a word is encountered in and which may relate to how context-dependent versus context-independent semantic access is for a given word (Hoffman, Lambon Ralph, & Rogers, 2013; additional analyses also ruled out confounds due to other psycholinguistic variables such as familiarity, sense dominance, and age of acquisition). Furthermore, they discover convergent support for these results in a prior study by Tokowicz and Kroll (2007) that reported effects of translation ambiguity on less concrete but not more concrete words. Re-analyzing the Tokowicz and Kroll items, Jager and colleagues found that they were primarily comprised of polysemes, thus the overall ambiguity effects are likely more specifically polysemy effects—a pattern that is particularly apparent in Tokowicz and Kroll’s monolingual lexical decision experiment. To explain these results, Jager and Cleland (2014) theorized that the breadth of the attractor basin formed by the neural network that encodes the semantic representation of more concrete words is much broader than that of the less concrete words, which reduces the degree of coherence (cooperation) that may occur among senses.

The report that polysemy interacts with concreteness is intriguing, and opens up the possibility that other semantic ambiguity effects, such as those associated with homonymy, may also be shaped by this or other semantic properties. Before devoting extensive efforts to incorporating this finding into a mechanistic theory, however, additional study is likely warranted given some of the detailed aspects of the Jager and Cleland (2014) finding, and how their results compare with those from other studies. For instance, the authors note that the Rodd et al. (2002, which was also replicated by Beretta et al., 2005) study included a range of concreteness values that were not explicitly restricted as part of item selection, whereas Experiment 1 of Jager and Cleland study contained only more concrete items and Experiment 2 contained only less concrete items. In the case of the more concrete items, there was only a non-significant 3 ms advantage for the polysemes, whereas this advantage was significant and 15 ms in magnitude in the case of the less concrete items. However, the advantage is actually smaller than would be predicted if the polysemy advantage reported Rodd et al. (2002) and
Beretta et al. (2005), which ranged from 14 to 20 ms, was actually due to only a smaller subset of their items.

Of course, other psycholinguistic variables that differ between these investigations, such as frequency and length, may be further shaping these results and contribute to the discrepancy. However, this is true not only across these three articles, but also within the two experiments reported by Jager and Cleland (2014). Indeed, across the two experiments, although the unambiguous controls and polysemes were very well matched on a range of psycholinguistic properties, the mean value of several of these properties differed substantially between the two experiments. For instance, the mean word frequency in Experiment 1 was 64, whereas in Experiment 2 it was almost double that—114. Similarly, each word had only 4.8 neighbors in Experiment 1, whereas words had 7.3 neighbors in Experiment 2; and increase of 50%. Collectively, these differences leave open the possibility that it is not concreteness that is mediating the polysemy effect, but an interaction with other psycholinguistic properties (see Armstrong, Watson, & Plaut, 2012, for a related discussion of how mean differences in other potential confounds may have shaped the interpretation of age-of-acquisition and word frequency effects, as well as an approach to item selection that can help avoid these issues). The possibility that such an interaction occurred is also consistent with the mean latencies reported in each task, which were essentially identical for the words (547 in Experiment 1 vs. 541 in Experiment 2; the nonword latencies were note reported), despite the fact that word frequency and neighborhood size, both of which typically show facilitative effects in lexical decision, were much larger for the items in Experiment 2.

More broadly, the inclusion of items with greater or lesser concreteness across experiments may have generated effects analogous to how semantic activity can be modulated by blocking versus mixing abstract and concrete words, or by order effects due to the type of word that appeared in the first block of an experiment. In a study investigating these issues, Tolentino and Tokowicz (2009) observed that N400 magnitudes were more significantly negative for concrete words in all conditions except for the case in which the first block presented to participants contained concrete items. The authors suggested that the concreteness of the stimuli provided a context that helped bias the system to engage in either coarser or more fine-grained semantic processing—more concrete items bias the comprehension system to activate more specific, fine-grained semantic features, whereas more abstract items bias the system to activate less specific, coarser semantic representations. Thus, the interaction between concreteness and polysemy may, analogously, be due to the modulation of semantic processing by similar context effects (see also Frisson, 2015).

Additionally, the consistent report of stronger polysemy effects for metonymic polysemes with more overlap between interpretations relative to metaphoric polysemes with overlap across interpretations (e.g., Klepousniotou & Baum, 2007; Klepousniotou, Titone, & Romero, 2008) also must be taken into consideration when contrasting the low- versus high-contrast polysemes. These points notwithstanding, the possibilities raised above do not imply that the conclusions drawn by Jager and Cleland (2014) are false. Rather, they indicate that these recent findings warrant further investigation and replication in other controlled settings, and may be useful conditions for other researchers to include as part of their stimulus selection and/or analyses, thus providing clear and robust insight into how concreteness and polysemy should be treated in a general theory of ambiguity resolution.

To recap, the results reviewed thus far suggest the following preliminary conclusions: Individual lexical decision studies have, in some cases, disagreed with respect to the relative differences between homonymous, polysemous, and unambiguous items. They have, nevertheless, generally shown at least a consistent rank ordering of these word classes, such that polysemous items were responded to most quickly, followed by the unambiguous controls, and lastly by the homonymous items. Careful re-interpretation of the results of these studies also suggests that visual lexical decision is generally associated with fast responses and a polysemy advantage in the absence of any effect of homonymy. In contrast, auditory lexical decision is generally associated with slower responses and both a polysemy advantage and a homonymy disadvantage.

2.1.5 Studies in Japanese. Contradicting these preliminary conclusions, however, are a series of studies conducted by Hino et al. (2006, 2010). All of this work was conducted in Japanese, which offered some unique possibilities in terms of nonword foil manipulations in examining ambiguity effects, but which also raises questions regarding the generalizability of these results to single-script alphabetic languages such as English. In particular, Japanese employs three scripts, two of which were the main focus of these investigations: Katakana, which is analogous to the alphabetic/phonetic representational structure of written English; and Kanji, in which individual characters each correspond to separate morphemes, and, to a first approximation, syllables. In one visual lexical decision study, Hino et al. (2006, Expt. 1) reported faster performance for words with less related interpretations (homonyms) and for words with more related interpretations (polysemes) than for unambiguous controls—that is, an overall ambiguity advantage. A similar overall ambiguity advantage has also been reported by Pexman and Lupker (1999) and by Hargreaves, Pexman, Pittman, and Goodeyer (2011) although in neither of those studies were the results significant by items.3 Lin and Ahrens (2010) also

---

3 The ambiguous items in the Hargreaves et al. (2011) study did show a processing advantage when the corresponding latencies and accuracies for their items were examined in the eLexicon database.
reported an overall ambiguity advantage based on subjective assessments of the senses and meanings associated with a word; however, the effect may, in fact, be a polysemy advantage as their regression analyses did not include number of senses as a covariate.

Hino et al. (2010) replicated and extended their original work. Once more, they observed an overall ambiguity advantage using Katakana nonwords in a task very similar to that used by Hino et al. (2006, Expt. 1). They then added Kanji words and nonwords as fillers to their item set in an effort to alter the difficulty of the task (Expt. 2). The Kanji words were known pairings of semantically related Kanji characters, whereas the Kanji nonwords were unknown pairings of semantically unrelated characters. The addition of these fillers altered the pattern of results substantially, such that only a polysemy advantage was observed. Finally, they re-introduced the overall ambiguity advantage by replacing the Kanji nonwords from the second experiment with Kanji nonwords formed by transposing the characters in a known pairing of semantically related Kanji characters into an unfamiliar order (Expt. 3). Strikingly, this modulation of nonword difficulty—and, consequently, of overall response latencies—had a non-monotonic effect on the presence of a homonymy disadvantage.

Collectively, the results obtained in the aforementioned studies conducted in Japanese appear to contradict those of several similar studies. However, even setting aside the complexities that arise in the contributions from multiple highly-different scripts, several important differences between these and other experiments may contribute to these discrepancies. First, the overall ambiguity advantage reported in several of the Japanese experiments may be due to how the different word classes have been labeled and contrasted by different groups of experimenters. For instance, in the Hino et al. (2006, 2010) studies, the unambiguous items had fewer interpretations than both the homonymous and polysemous items. These items were then compared to polysemes and homonyms that had approximately equal numbers of interpretations (~5) and differed only in whether these senses were loaded onto a single meaning or distributed across multiple meanings. Using the definitions adopted in this review, the homonymous items used by Hino and colleagues could thus alternatively be classified as hybrid items that are both homonymous and polysemous, given that these items have more senses than the unambiguous controls. In contrast, studies that have used (relatively) unambiguous controls that have the same small total number of (related) interpretations as homonyms have often generated a homonymy disadvantage (e.g., Beretta et al., 2005; Rodd et al., 2002), whereas hybrid items have shown at least a numeric processing advantage (Armstrong & Plaut, 2011).

The results reported by Klepousniotou and Baum (2007) in two lexical decision studies lend further support to this distinction between item classes. They defined unambiguous words as having a single interpretation, homonymous words as having two unrelated interpretations, and polysemous words as having two related interpretations. Using these definitions, they observed the standard polysemy advantage and a numeric homonymy disadvantage relative to unambiguous controls.

Admittedly, differences in definition can only partially explain the discrepant findings reported in Japanese across all types of nonword fillers. In particular, the findings of a homonymy advantage should receive further investigation to establish their reliability and generalizability. This is particularly true with respect to the complex and non-monotonic modulation of the presence of a homonymy disadvantage as a function of increased nonword difficulty in the Hino et al. (2010) study, in which only 14 target homonyms were tested.

2.1.6 Semantic categorization and relatedness judgment. Moving away from lexical decision, several researchers have employed tasks that require explicit judgments related to the interpretations of a word. These studies have the apparent advantage of involving semantic processing directly, in contrast to lexical decision, which is often thought to be based on “lexical” rather than semantic representations. Two particularly popular tasks in this literature are semantic categorization (e.g., does a word refer to a living thing) and meaning relatedness judgments. These tasks have typically resulted in a strong homonymy disadvantage and slower overall latencies than lexical decision (e.g., Gottlob, Goldinger, Stone, & Van Orden, 1999; Piercey & Joordens, 2000). However, unlike for lexical decision, in which all meanings provide evidence for the same response (“yes”), interpreting the results of explicit semantic tasks requires careful consideration of how each interpretation of a word relates to each possible response. For instance, Pexman, Hino, and Lupker (2004) demonstrated that at least some of the reported homonymy disadvantage effects in these paradigms are due to response competition because not all of the meanings of a homonym fell into a target category or were associated with a probe word. For example, BANK has two interpretations (related to <river> and <money>) that do not denote living things, whereas BAT has a <mammal> interpretation that denotes a living thing and a <baseball> interpretation that does not. When there was inconsistent support for a particular response across all of the meanings of a homonym—for instance, a “no” response for “is <mammal>/<baseball> BAT a living thing?”—a homonymy disadvantage was detected, whereas when there was consistent support for a particular response—for instance, “is <river>/<money> BANK a living thing?”—the homonymy disadvantage disappeared. Paradoxically then, such explicit (Balota et al., 2007). However, a similar analysis of a larger set of homonyms failed to replicate that finding (Armstrong, Tokowicz, & Plaut, 2012).
semantic tasks may be even more at risk from post-semantic processing (for related discussion, see, e.g., Balota & Paul, 1996).

Hino et al. (2006; see also Hino, Lupker, & Pexman, 2002) have, nevertheless, established that a homonymy disadvantage can be observed under some circumstances and that it cannot be attributed to response competition, as discussed above (cf. Pexman et al., 2004). To do so, they conducted a series of semantic categorization experiments using the same set of Japanese word stimuli that produced an overall ambiguity advantage in lexical decision. In each semantic categorization experiment, participants were asked to make judgments based on a semantic category that varied from being defined either very narrowly or very broadly (e.g., “does the word refer to a vegetable?” vs. “does the word refer to a living thing?”). In all cases, all meanings of the homonyms were associated with a “no” response. The authors found that when broad categories were used, there was a significant processing disadvantage for homonyms (which might be more appropriately labeled as hybrid items, per the earlier discussion). Additionally, overall latencies in the experiment were several hundred milliseconds slower than in their lexical decision experiment. In the case of narrow categories, however, responses were extremely fast—faster, in fact, than in the lexical decision experiment—and no effects of ambiguity were observed (see also Armstrong, 2007; Hargreaves et al., 2011; Pexman & Lupker, 1999; Siakaluk, Pexman, Sears, & Owen, 2007, for similar results in semantic categorization even when overall latencies are somewhat longer).

Interestingly, similar null or substantially reduced effects of ambiguity have also been reported in lexical decision studies when relatively easy nonwords were employed (e.g., Azuma & Van Orden, 1997; Borowsky & Masson, 1996; Rodd et al., 2002). Consequently, it is possible that the reduction of ambiguity effects observed in narrow semantic categorization is due to the fact that only a relatively coarse semantic representation could be activated sufficiently early to influence performance, and that such a representation does not differ greatly between ambiguous and unambiguous words.

A study by Siakaluk et al. (2007) further supports the notion that processing time is a relevant factor in understanding semantic ambiguity effects in categorization judgment tasks. They found that the standard yes/no version of semantic categorization involving narrow categories such as “animal” did not yield homonymy disadvantages, but a slower go/no-go version of the task did give rise to such effects. Note that these tasks involve only very minor changes to the response system itself: in model fits of parameters related to the underlying evidence in behavioral experiments, the “evidence accumulation parameter” (e.g., drift rate) remains the same, and only the threshold at which a response is generated changes, allowing additional evidence to accumulate before responding (Gomez, Ratcliff, & Perea, 2007; Shenoy & Yu, 2012; Siakaluk, Buchanan, & Westbury, 2003; see also supporting behavioral evidence from Perea, Rosa, & Gómez, 2002). Thus, the different patterns of behavior results likely reflect how information within the lexical system changes over time rather than qualitative changes to a post-lexical response system.

Contributing additional insight into semantic categorization data are several studies related to their neural bases. In an fMRI study, Hargreaves et al. (2011) found that ambiguous words produced significantly greater activation in the left inferior frontal gyrus—which, in their view, provides inhibition to semantics when incompatible meanings of an ambiguous word are activated—even in the context of narrow semantic categories that do not show any behavioral ambiguity effects (see also Bekinschtein, Davis, Rodd, & Owen, 2011). This view and interpretation of the fMRI data are, however, not without controversy. For instance, Rodd, Johnsrude, and Davis (2010) used a task-interference paradigm and concluded that the computations related to ambiguous word processing in the left frontal gyrus occur only several seconds after the ambiguous stimulus is initially processed. Thus, although relevant to understanding late ambiguity effects or the re-evaluation of different incompatible meanings, the effects reported by Hargreaves et al. (2011) may not be modulating early neural processing dynamics, which are occurring 250-650 ms following stimulus presentation (based on ERP and MEG data). Consequently, the fMRI dynamics that they report may not be driving performance in the behavioral semantic categorization tasks, in which latencies typically fall within this early time-frame.

More recent data reported by Rodd, Johnsrude, and Davis (2012) using a fast-acquisition fMRI procedure has, however, supported the claim of left inferior frontal gyrus involvement in the immediate processing dynamics of an ambiguous word both before and after context. However, they suggest that this brain region is involved in the reinterpretation of an incorrectly interpreted ambiguous word. More work is needed to fully flesh out the role of inferior frontal regions and the time-course of their contributions—particularly with complementary techniques with improved temporal resolution such as MEG and EEG. However, the studies at least agree that, in addition to the anterior temporal lobes, frontal regions are also making important contributions to processing semantically ambiguous words, and are involved in the meaning selection process at least on a long time-scale of 500 ms to multiple seconds.

In summary, in the absence of biasing context, visual lexical decision tasks tend to show an early polysemy advantage and at best a very weak homonymy disadvantage. Auditory lexical decision tasks produce similar results but with more reliable homonymy effects and with considerably longer overall latencies. Semantic categorization and relat-
edness judgment tasks that involve broad categories and that avoid response competition confounds produce considerably longer latencies than comparable visual lexical decision tasks and a homonymy disadvantage. Similar tasks that involve narrow categories produce very fast responses and, as in similarly rapid lexical decision tasks, no ambiguity effects whatsoever. The neuroimaging data support a neural basis for these effects in the anterior temporal lobe, which has long been associated with semantic processing, as well as relevant contributions from the (primarily left) inferior frontal gyrus. Importantly, the neuroimaging data also fail to identify a basis for these effects in more posterior brain regions typically associated with the early orthographic and phonological processing of a word.

2.2 Ambiguous words in biasing contexts

A range of tasks have been employed to study how context shapes the activation of a particular interpretation of an ambiguous word, from single-word priming (e.g., MONEYBANK) to naturalistic reading. Although there is no general agreement of exactly how a context biases processing (but see McClelland et al., 1989 for one account), there is agreement in subjective assessments of what constitutes a biasing context (Paul, Kellas, Martin, & Clark, 1992). Moreover, the effects of contextual bias generally cannot be explained by simple lexical association (Simpson, 1994), contrary to one previous theory (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982), and context has a similar impact on processing regardless of whether it provides a local (e.g., within a sentence) or global (e.g., within a paragraph) constraint, so long as the two constraints do not disagree (Kambe, Rayner, & Duffy, 2001).

2.2.1 Limited early effects of context. In a classic study, Swinney (1979; see also Conrad, 1974; Kintsch & Mross, 1985; Oden & Spira, 1983; Onifer & Swinney, 1981; Seidenberg et al., 1982) tested visual lexical decision following a homonym in a spoken sentence context that was biased toward one of its meanings. He found that, at a short SOA, target words related to any of the meanings of the homonym were primed, whereas at a longer SOA, priming was observed only for words related to the contextually-appropriate meaning. Performance for the inconsistent meaning was not, however, any slower than that for an unrelated meaning, indicating that the inconsistent interpretation was not suppressed below baseline. A similar reduction or elimination of priming for the inconsistent meaning of a homonym was also obtained in spaced-presentation repetition effects of same versus different meanings in lexical decision and naming when the inter-stimulus interval was brief (Masson & Freedman, 1990; Van Petten & Kutas, 1987). The original Swinney (1979) study was followed up by a similar study that focused on polysemes (Williams, 1992). The results of that study also showed early activation of both the consistent and inconsistent senses. In contrast to the studies with homonyms, however, the facilitation for both senses persisted at longer SOAs, although to a numerically reduced extent for the inconsistent sense.

Van Petten and Kutas (1987) conducted a similar experiment to that of Swinney (1979) but recorded ERP waveforms and did not require participants to make any explicit responses. Their study thus eliminated any decision-making element of the task and might therefore be a purer measure of semantic processing dynamics. They found that N400 amplitudes recapitulated the behavioral effects and therefore could be the neural source that drives them. Specifically, at a long (700 ms) SOA, words related to the inconsistent meaning showed a similar large N400 amplitude as for unrelated words. In contrast, at a short (200 ms) SOA the inconsistent items, which showed significantly larger N400s than the consistent items, nevertheless elicited a smaller N400 relative to the unrelated words. Interpreted within the classic view of the N400 as measuring ease of semantic integration (Kutas & Hillyard, 1980; see Kutas & Federmeier, 2011; Laszlo & Plaut, 2011, 2012; Laszlo & Armstrong, 2014), these effects suggest that the effect of prior context is limited early in the timecourse of processing a word.

Recent work by Klepousniotou et al. (2008), however, has shown that the effect of context may be earlier than reported in the van Petten and Kutas (1987) study. In this work, participants completed a single-word primed visual lexical decision task in which the prime was either unrelated to a target word, or related to the subordinate or dominant interpretation of balanced and unbalanced homonyms and metonymic polysemes. The prime preceded the target by 250 ms; thus, approximately the same amount of time was available for context to bias processing as in the prior studies by and Swinney (1979) and Van Petten and Kutas (1987). Overall, Klepousniotou and colleagues observed significant effects only between the word classes in the N400 component, suggesting a semantic basis for the effects of context and not an effect related to form-based priming. More specifically, they reported reduced N400 amplitudes when a prime related to either of the interpretations of an ambiguous word was presented. This reduction interacted both with whether the dominant or subordinate interpretation was primed and the degree of relatedness among the ambiguous word’s interpretations. Greater reductions in N400 amplitude were observed for the dominant interpretations relative to the subordinate interpretations of unbalanced homonyms, and this differential reduction decreased as the relatedness among interpretations increased and when the frequencies of the subordinate and dominant meanings were relatively balanced. These results further suggest that semantic integration has begun but has not been completed even when the context precedes the target by 250 ms (cf. Swinney, 1979), that the word form more strongly activates the dominant interpretation of...
a word, and that more strongly related interpretations share overlapping semantic representations.

### 2.2.2 Strongly biasing context

Given the differences in the time-course of context effects reported in the prior studies, one important question that has been raised is whether very early processing is necessarily contextually independent (Swinney, 1979), or whether the absence of a very early effect of context in these studies was due to the use of relatively weak and insufficiently biasing contexts. Several studies have investigated these two possible alternatives. In general, these studies are in agreement that in strongly biasing circumstances, at least small effects of context can be detected during very early processing. For instance, Klepousniotou (2002) reported the results of a lexical decision task in which legal nonword foils and target homonymous, metaphorically polysemous, and metonymically polysemous words were first primed by sentences consistent with either the dominant or subordinate interpretation of the ambiguous words. She found that, relative to unrelated unambiguous controls, all of the target ambiguous words showed a processing advantage from context. This indicates that context exerts a small but significant effect on early processing.\(^4\) Seidenberg et al. (1982) also reported similar early contextual biases and found that the magnitude of the processing advantage increased as semantic overlap between the interpretations increased, paralleling the processing advantage as a function of semantic overlap observed in later work (e.g., Klepousniotou & Baum, 2007).

Pushing the issue further, several additional studies have employed extremely biasing context and observed that the inconsistent meaning of a homonym may not be activated at all after the first 120-140 ms of processing of a target (Martin, Vu, Kellas, & Metcalf, 1999; Sheridan & Reingold, 2012; Simpson, 1981; Tabossi, 1988). This strong bias was more consistently demonstrated for the suppression of a subordinate meaning in a dominant context than in a reverse situation (Forster & Bednall, 1976; Hogaboam & Perfetti, 1975; Neill, Hilliard, & Cooper, 1988). Similar patterns have also been observed in the explicit ambiguity judgment tasks reported by Simpson (1981), and in naturalistic reading (but not sensicality judgments) in the case of polysemes (Frisson, 2015). Collectively, these results support a graded view of contextual influences and discount earlier staged context-free then context-sensitive views (Swinney, 1979), and also support a shared-features theory of the representation of polysemous senses. However, they are still generally consistent with the notion of early processing being relatively context-independent and later processing being more context-sensitive, with the strength of the contextual bias determining the time at which this graded transition takes place (cf. theories that predict strong amounts of constraint from highly predictive contexts, e.g., Altmann & Mirković, 2009).

Moreover, the consistent finding that the contextually inappropriate interpretation of an ambiguous word is not responded to more slowly than an unrelated control item provides important insight into the nature of the inhibitory processes that suppress contextually-inconsistent interpretations. Specifically, the results reviewed so far suggest that a general inhibitory process is suppressing overall activation of inconsistent semantic representations, consistent with the proposed role of inhibition as a non-specific activation regulation mechanism (Hargreaves et al., 2011), and not as a highly specific processes that is selectively targeting the contextually inappropriate interpretation. We return to this point in the discussion.

### 2.2.3 Patient populations

Insights into context-sensitive ambiguous word comprehension have also been garnered from patient populations, including Broca’s aphasics (Swaab, Brown, & Hagoort, 1998), mild cognitive impairment (Taler et al., 2009), and Alzheimer’s-type dementia (Balota & Ducheck, 1991; Faust, Balota, Ducheck, Gernsbacher, & Smith, 1997). In general, these patients show inhibitory effects for inconsistent meanings and facilitatory effects for the consistent meanings that are both weaker and slower than control participants. Reduced inhibition has also been reported to a lesser extent in older controls (Dagerman, MacDonald, & Harm, 2006). These findings, in association with the fact that these patients often suffer from the differential impairment of a particular brain region (e.g., Broca’s area in frontal cortex) provide convergent evidence for at least a partial separation of inhibitory and excitatory mechanisms related to ambiguous word comprehension. Nevertheless, the results of these studies should be interpreted cautiously because evidence from other tasks, such as relatedness judgments, calls into question whether children (e.g., Booth, Harasaki, & Burman, 2006), and presumably, adults with neurological impairments, attend to the context properly (Khanna & Boland, 2010), or whether limited working memory capacity may restrict the influence of context (Khanna & Boland, 2010, but see Mason & Just, 2007, for evidence that increased working memory is associated with longer-lasting as opposed to shorter-lasting activation of a word’s multiple interpretations).

### 2.2.4 Naturalistic tasks

Moving beyond highly-controlled tasks with low ecological validity, several studies have investigated the effects of ambiguity in more realistic settings such as self-paced reading. An additional advance...
tage of this approach is that it minimizes the task and response demands that complicate the interpretation of many studies (e.g., as illustrated by Pexman et al., 2004). Frazier and Rayner (1990) found evidence supporting a semantic basis for the effects of context on homonym processing in a naturalistic reading paradigm that involved measuring eye-fixation duration via eye-tracking (see also Duffy, Morris, & Rayner, 1988; see Frisson & Pickering, 1999, for a follow-up focusing on different types of polysemes). They examined participants’ gaze durations for homonyms and polysemes relative to unambiguous controls presented either before or after a biasing context. The results indicated that, following the presentation of a biasing context, participants’ gaze durations were longer for all ambiguous words relative to the unambiguous controls. This suggests that fully resolving a semantic ambiguity impedes processing to some extent for all ambiguous words. Similarly, Piercey and Joordens (2000) also observed a processing disadvantage for polysemes when contextual integration must take place to generate a response. However, when disambiguating context was not presented before the ambiguous words, responses were slower for the homonyms only, suggesting that the presence of multiple meanings (as opposed to multiple senses) is, in and of itself, sufficient to generate a processing disadvantage. Recent work by Brocher, Foraker, and Koenig (2012) has also provided even finer-grained insight into these dynamics, showing that context integration appears to occur more rapidly in the case of polysemes than homonyms, and is faster when the interpretations of a polyseme are strongly related to one another (see Binder & Rayner, 1998, 1999; Frisson & Pickering, 1999; Foraker & Murphy, 2012, for additional data consistent with this perspective after dividing a mixed group of ambiguous words into homonyms and polysemes; cf. Kellas & Vu, 1999). Insofar as eye movements are initiated once the bulk of the processing of the perceived word has been completed (see, e.g., Reichle, Pollatsek, Fisher, & Rayner, 1998), these findings dovetail with the previously discussed observation of a homonym disadvantage in the absence of a strongly biasing context in the semantic categorization tasks involving broad categories (e.g., Hino et al., 2006). A natural interpretation of these data, therefore, is that both semantic categorization and naturalistic reading activate and are influenced by semantic representations. Similarly, in the presence of a biasing context, the detection of a processing disadvantage for all ambiguous words that was larger for words with less related interpretations is consistent with a graded degree of competition between the interpretations as part of the semantic access and contextual integration processes. Moreover, the fixation durations, which were typically less than 300 ms, generally agree with the latencies of N400 and M350 ambiguity effects reported using EEG and MEG (Beretta et al., 2005; Laszlo & Federman, 2011; Taler et al., 2009; Van Petten & Kutas, 1987). This congruency, as well as prior arguments supporting the direct comparison of neurophysiological measures and eye-tracking results (e.g., Sereno, Rayner, & Posner, 1998) further supports a semantic locus for the reported semantic ambiguity effects.

2.2.5 Representations of ambiguous words. Other work has focused on the nature of the representations of ambiguous words themselves and how they interact with comprehension processes. Klein and Murphy (2001, 2002) attempted to determine whether multiple related senses overlap and share a substantial “core” interpretation, as had been assumed by some other researchers (e.g., Frazier & Rayner, 1990; Rodd et al., 2002, 2004), or whether each sense was represented essentially separately, as is generally assumed for the different meanings of homonyms. Klein and Murphy (2001) presented a polyseme twice across two phrases that tapped either the same sense or a different sense of the word (e.g., PAPER as a material or as a publication printed on that material; see also Bainbridge, Lewandowsky, & Kirsner, 1993). They found that presenting a second sentence using the same sense of the polyseme improved recall of the polyseme in a later memory test and also decreased latencies in a sensicality judgment task, relative to sentences involving different or neutral/nonsensical senses. Moreover, the latency differences between the same- and different-prime conditions in the sensicality judgment were only slightly smaller for polysemes than for homonyms. Further, polysemes showed facilitation for consistent senses and inhibition for inconsistent senses relative to a neutral baseline (see also the similar effects that were reported by Seidenberg et al., 1982). Similar results were obtained in Klein and Murphy’s (2002) follow-up studies, in which they reported that phrases using different senses of a polyseme were only rarely categorized together in the presence of a thematic lure in a forced-classification task. The low rates of grouping the polysemes together were only slightly (but significantly) higher than those of grouping the different meanings of the homonyms together.

Based on their results, Klein and Murphy (2001, 2002) argued that there is only minimal semantic overlap across related senses, and so relatedness must be encoded separately from semantic overlap (see also the lexical decision data from Hino et al., 2010). There are, however, several facts that challenge this conclusion. From the outset, Klein and Murphy note that, although they were studying polysemes with related senses, the items that they employed were “clearly distinct in meaning” (Klein & Murphy, 2002, p. 568). This
could obviously be directly responsible for the failure to find strong evidence for featural overlap and for differences between homonymy and polysemy. Additionally, Pykkänen, Linna, and Murphy (2006) replicated the behavioral task employed in the Klein and Murphy (2002) study while recording MEG data. In contrast to the original behavioral findings, they found earlier M350 latencies for polysemes and later M350 latencies for homonyms, particularly in the left hemisphere. These results suggest that Klein and Murphy’s behavioral task and associated measures may not be sufficiently sensitive to the semantic processing that occurs during these tasks (see also Beretta et al., 2005). Pykkänen and colleagues argued that, whereas homonyms are represented separately, the different senses of a polyseme do share a core interpretation.

Klepousniotou et al. (2008; see also Williams, 1992) also failed to reproduce the same patterns of effects reported by Klein and Murphy (2001, 2002) when using a finer-grained break-down of the relatedness continuum into low-, medium- and high-overlap items—roughly equivalent to homonyms, metaphoric polysemes, and metonymic polysemes. In contrast to Klein and Murphy’s results, Klepousniotou and colleagues found that the latency difference across phrases priming two different interpretations of an ambiguous word were relatively small for the high-overlap items. Similar to Klein and Murphy’s results, however, they did find a large latency difference across phrases priming two different interpretations of the medium- and low-overlap items. They also failed to find a difference in performance between the medium- and low-overlap items, suggesting that medium-overlap items may be more similar to low-overlap items than high-overlap items, given their coding scheme. Regrettably, their unambiguous control condition was not successful at establishing baseline performance in an indisputable manner (see Klein & Murphy, 2001; Klepousniotou et al., 2008; Plaut & Booth, 2000, for discussion of different baselining methods). Consequently, it remains unclear whether all of these effects should be viewed exclusively as competitive slow-downs consistent with an overall ambiguity disadvantage, or whether a portion of these results should be viewed as enhanced processing relative to a neutral baseline. Nevertheless, these results suggest substantial commonality in the representation of the high-overlap items, but this commonality decreases rapidly between the high- and medium-overlap items. Consistent with this interpretation, Klepousniotou et al. (2008) found that most of the polysemous items used by Klein and Murphy would be considered to have either low- or medium-overlap according to their classification scheme.

In summary, the implications of studies of semantic ambiguity involving biasing context are as follows. In terms of representations, polysemes with highly related interpretations produce effects consistent with their having more shared features than homonymous words. In terms of processing dynamics, early context effects are typically weak, and multiple interpretations are activated. As the influence of context increases, the contextually-appropriate interpretation remains active but the inappropriate interpretation’s activity decreases as a function of its relatedness to the appropriate interpretation. The timecourse of de-activation of the inappropriate interpretation varies as a function of contextual constraint and interpretation dominance. This de-activation generally appears to drop the activation of the inappropriate interpretation to the same level as that established by neutral control words; if any selective de-activation does occur, its effects are therefore quite weak, at best. Developmental, clinical, and neuroimaging data provide support for these effects being caused by processing dynamics in the anterior temporal lobe regions that have a long history of being associated with semantic processing, and for the separation—both functional and neuroanatomical—of some inhibitory processes in frontal regions from the main semantic hub in the anterior temporal lobe.

3. Theoretical Accounts of Semantic Ambiguity Effects

Several influential theoretical accounts have been proposed to explain various aspects of the reviewed findings. Given the extensive history of this literature, we restrict our consideration to more general accounts, excluding those that are relatively task-specific (e.g., Van Petten & Kutas, 1987). An exception was made, however, for theories related to lexical decision, given the prominent role this task has played in semantic ambiguity research and word recognition research more generally. The different accounts are grouped based on whether the locus of ambiguity effects is believed to be in a word’s surface representations (orthography and/or phonology), in its semantic representations, or in the response selection (decision-making) system.

3.1 Orthographic/phonological accounts

There is a long history of theoretical accounts of visual lexical decision performance being based on the orthographic—as opposed to the semantic—representation of a word, either primarily (Balota & Chumbley, 1984; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2000; Grainger & Jacobs, 1996; Hino & Lupker, 1996; Hino et al., 2002; Kawamoto, 1993; Kawamoto, Farrar, & Kello, 1994; Pexman & Lupker, 1999), or at least partially (Seidenberg & Vigneau, 2012; Saffran, 2015).

Hino et al. (2010) have argued that Klepousniotou et al.’s (2008) high-overlap items should properly be considered unambiguous (see also Hino et al., 2006, p. 182), but this position is undermined by previous findings that such high-overlap polysemes produce different effects than true unambiguous words with only a single sense (Klepousniotou & Baum, 2007). Moreover, recent work by Rodd, Cutrin, Kirsch, Millar, and Davis (2013) supports Klepousniotou and Baum’s use of lexicographic definitions to determine semantic feature overlap.
McClelland, 1989). For instance, Kawamoto (1993; see also Kawamoto et al., 1994) developed a connectionist simulation of how the time-course of orthographic feature activation—hereafter referred to as settling—could account for the overall ambiguity advantage that was the accepted pattern of results at the time. According to his account, ambiguous words receive less consistent feedback from semantics because different semantic features are activated for the different meanings. As a result, ambiguous words develop stronger mappings within orthography when the network is trained to activate the orthographic and semantic representations of a word, whereas unambiguous words rely more on the consistent feedback from semantics and develop weaker intra-orthographic mappings. Building on a similar theme, Hino and Lupker (1996; see also Hino et al., 2010, 2002, 2006; Pexman et al., 2004; Pexman & Lupker, 1999) suggested that, when an ambiguous word engages semantics, it will generate more semantic activation than an unambiguous word. This activation will then feed back and more strongly activate the ambiguous word’s orthography, leading to an ambiguity advantage. Support for this prediction is also provided in the form of slower lexical decision latencies for words that have high-frequency synonyms (Hino et al., 2002).

Relatedly, but with effects more in line with current findings, Seidenberg and McClelland (1989) reported higher error rates for homographs (e.g., <breeze>/<to turn> WIND) than non-homographs in the phonological representations of model simulating how orthography maps to phonology. This was theorized to be due to weaker and less consistent nature of the orthographic-to-phonological mappings for the homographs, and correlates with reports of slower processing for homographs than for non-homographs (Seidenberg, Waters, Barnes, & Tanenhaus, 1984). Similar work by Newman and Joanisse (2011) with homophones also supports this general position, and points to a later frontal contribution involved in longer and more complete settling.

There are a number of reasons that orthogonal (or phonological) accounts of semantic ambiguity effects are less than fully satisfactory, beyond the fact that some are inconsistent with current empirical findings. In lexical decision, orthography alone does not provide a basis for accurate responding when using well-matched nonwords, and so any differences between words and nonwords must result from semantic (or “lexical”) feedback. In this case, however, the response system might just as well monitor semantics (or the “lexicon”) directly, as this would enable more rapid responding (Armstrong & Plaut, 2011; Azuma & Van Orden, 1997; Rodd et al., 2002). Moreover, semantic ambiguity effects manifest in tasks or contexts that either lack orthographic input (e.g., auditory sentence contexts; Swinney, 1979) or are explicitly semantic (e.g., category verification, sensicality judgments). Although nothing precludes back-activation of orthography in these tasks, it is unclear why this would be relevant or useful. Insofar as semantics is relevant (and sometimes required) for task performance, it makes more sense to propose that responses are based on, or directly influenced by, semantic activation.

Thus, although orthographic activation may play a minor role in contributing to some semantic ambiguity effects, there are empirical and theoretical reasons to focus attention on downstream processes related to semantic activation and response selection.

3.2 Semantic accounts

Another common type of account of semantic ambiguity effects ascribes them to differences in featural richness, and the related notion of overall semantic activation. It is well established that featurally rich items, such as concrete nouns (e.g., TRUMPPET), are typically processed more rapidly than less rich items, such as abstract nouns (e.g., JUSTICE), presumably because concrete nouns are associated with more semantic features (Plaut & Shallice, 1993). Polysemes, which may activate large numbers of core features common to all of the word’s interpretations, have been hypothesized to produce an ambiguity advantage that is, in part, due to richness (Hargreaves et al., 2011; Klepousniotou & Baum, 2007; Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008; Pexman, Lupker, & Hino, 2002; Rodd et al., 2002). However, it is unclear why homonyms, which must be at least moderately rich when the total number of features across their multiple meanings are compared, do not generally show a similar processing advantage. Additionally, other studies that control for a measure of richness still yield significant ambiguity effects (e.g., Expt. 2 of Rodd et al., 2002). Thus, featural richness may be more appropriately treated as a potential confound in ambiguity experiments, as is the case with factors such as frequency and familiarity (Gernsbacher, 1984), rather than as an explanation of ambiguity effects per se.

A second type of semantic account has focused on the dynamics of settling within semantics for words with different numbers of interpretations, and different degrees of featural overlap across those meanings, to explain the typical lexical decision results (Rodd et al., 2004). This account was successful in predicting that polysemous words are processed more quickly than unambiguous words, which are in turn processed more quickly than homonymous words. According to this account, the polysemiese advantage arises because words with many related senses develop broader basins of attraction within semantics than those with few senses, whereas the homonymy disadvantage arises due to interference between the different meanings. An explicit connectionist implementation of this account also successfully substantiated these predictions, although it did predict a stronger homonymy disadvantage and a weaker polysemy advantage,
whereas the opposite is true in most of the behavioral data. This account does, however, suffer from the critical limitation of being explicitly cast as a model of semantic processing in lexical decision, and was not designed as a general model of comprehension per se. Consequently, this account would need to be extended to capture the data from other tasks.

Arguably, the semantics-based account that has captured the broadest set of different results was outlined by Piercey and Joordens (2000). They proposed that different tasks require different degrees of precision in the underlying semantic code for participants to generate fast and accurate responses. For instance, tasks such as isolated-word lexical decision typically require only the activation of an imprecise semantic representation because a “yes” response can be made if the stimulus evokes any interpretation(s) whatsoever; the full activation of the contextually-appropriate interpretation is not necessary. In such tasks, ambiguous words would be associated with an ambiguity advantage because more semantic features are active early in processing, before contextually-inconsistent interpretations are suppressed. In contrast, other tasks require a more precise semantic representation that is constrained by context. Activating such a precise semantic representation would take more time for ambiguous words because these words need to interact with context to activate an appropriate interpretation. Unambiguous words therefore could have their semantic representation resolved more quickly because precise semantic activation can be driven by the word form alone.

The Piercey and Joordens (2000) account has many of the qualities necessary for a general account of semantic ambiguity resolution, in that it is based on semantic processing and holds promise for accounting for a range of different effects. However, it suffers from the critical flaw that, in its current form, it does not account for the different effects of relatedness of meaning that have been reported in different tasks. For instance, it predicts an ambiguity advantage in lexical decision tasks for homonymous words, although both no homonymy effect or a homonymy disadvantage is typically observed. Similarly, the account predicts a processing disadvantage for polysemes in semantic categorization, whereas polysemes typically perform identically to unambiguous controls in that task. More refinement and elaboration of this theory, for instance, by combining it with a consideration of settling dynamics within semantics (Rodd et al., 2004), would therefore be needed for it to account for the currently-accepted effects in the literature. Indeed, our account can be seen as just such an elaboration.

3.3 Response selection accounts

An account based on task differences within the response selection system (e.g., Hino et al., 2006, 2010; Pexman et al., 2004) represents an interesting alternative to accounts based on the orthographic, phonological or semantic components of the word recognition system. This account assumes that the operation of the lexical system is more-or-less identical across tasks, and that therefore the same semantic representation is derived. Thus, differences in semantic ambiguity effects that are observed across tasks cannot be due to differences within the lexical system itself but must arise from differences in how responses are selected based on the semantic representation.

Initial support for a response-based account came from work from semantic categorization tasks in which the different meanings of a homonym may either provide partial support for both a “yes” and a “no” response, or support only one response (Pexman et al., 2004). Pexman and colleagues reported that the homonymy effects dissipated once response competition due to partial support of each response by different meanings was eliminated. In some circumstances, then, a response selection account is clearly a solid basis for explaining the observed effects.

Follow-up work has complicated matters, however. For instance, Hino et al. (2006) reported an ambiguity disadvantage in semantic categorization tasks involving broad categories (e.g., "living thing"), but not for narrow categories (e.g., "vegetable") even when all interpretations of a homonym support a “no” response. To explain these data, Hino and colleagues proposed that the decision system will differ, qualitatively, in what semantic features it attends to in different versions of the task—for instance, a few core features for a narrow category versus a broader, fuzzier set of features for a broad category.

This style of account appears to have the necessary explanatory power to deal with effectively with the relevant behavioral phenomena. Indeed, it is arguably too flexible in its current verbal formulation, because different tasks that generate new and different results can always be argued to differ qualitatively to some degree. Consequently, evaluating the validity of this theory requires that it be described in more explicit and mechanistic terms that specify how and why the decision system is configured in different ways in different tasks, and how it interacts with semantics and other parts of the lexical system. To date, however, no such formulation of the response system account has been reported in the literature. The degree to which a response system would actually behave as predicted and explain between-task variability, therefore, remains an open question. Moreover, even assuming that a more explicit version of the account could be developed, it would, nevertheless, still be at odds with neuroimaging findings that implicate a semantic basis for many of the observed effects. Collectively then, although considering the contributions from the response system to ambiguity effects has identified some potential confounds in past experimental designs, it seems unlikely to provide a full account of such effects.
4. The Semantic Settling Dynamics Account

Although a complete account of ambiguity effects will certainly involve some aspects of the reviewed accounts, as they stand those accounts are unsatisfactory in that they a) generate predictions that run contrary to currently-accepted patterns of effects in the literature, b) are relatively narrow in the scope of issues that they bear on, and/or c) are mechanistically underspecified and whether they operate as described verbally or can generate novel, testable predictions a priori is questionable. An improved account should address these limitations.

One basis for developing such an account was previously highlighted by Simpson (1994, p. 365): “it has become quite apparent that any explanation of ambiguity processes that does not include an account of changing activation patterns over time simply cannot hope to capture the complete picture of the processes relevant to the selection of a single appropriate meaning.” Along these lines, we introduce what we term the semantic settling dynamics (SSD) account that is developed within a more biologically-motivated variant of the standard connectionist framework (building upon our related work in this vein, see Armstrong, 2012; Armstrong & Plaut, 2008, 2011).

For ease of exposition, and consistent with our literature review, we first consider the nature of processing isolated words without biasing context, before considering how context modulates this processing. In addition, unless otherwise mentioned, the dynamics are described under the assumption that the relative frequencies of an ambiguous word’s interpretations are balanced. Insofar as this is not the case, the properties of ambiguous words are expected to regress back toward those of unambiguous words. Also, for simplicity, each interpretation is assumed to be equally semantically rich and be associated with the same number of semantic features. Finally, although we describe unambiguous, polysemous, and homonymous words as separate categories, it should be kept in mind that an extremely high degree of semantic overlap among senses will cause a polysemous to resemble an unambiguous word, whereas an extremely low amount of low overlap will cause it to resemble a homonym.

4.1 Processing ambiguous words in isolation

According to the account, the dynamics in semantics are due to the interaction between two factors: 1) representational differences between unambiguous, homonymous, and polysemous words, and 2) the architectural constraints and processing characteristics of the underlying neurocomputational system. The most critical characteristics of the computational system are that early processing is dominated by excitatory/cooperative dynamics, whereas later processing is dominated by inhibitory/competitive dynamics.

Several properties of individual neurons and neural systems give rise to these dynamics (for a general review, see Carreiras, Armstrong, Perea, & Frost, 2013). First, most projections between cortical regions are excitatory, whereas connections within a region are both excitatory and inhibitory. This contributes to a transient rise in excitation prior to the engagement of the local inhibitory circuit when a brain region receives new input. Moreover, compared to excitatory neurons, inhibitory neurons are generally fewer in number and have relatively non-specific projections, suggesting that they serve more to regulate overall activity among the excitatory neurons rather than communicate specific information themselves (for an explicit demonstration of this principle, see Laszlo & Plaut, 2012; Laszlo & Armstrong, 2014; for supporting behavioral evidence and related theories, see Becker, 1980; Khanna & Boland, 2010; Neely, 1991; O’Reilly & Munakata, 2000). A sub-population of inhibitory neurons may also have inherently slower/weaker activation dynamics because of its cellular characteristics, which would further exacerbate the differences between excitatory and inhibitory dynamics (Laszlo & Plaut, 2012; Laszlo & Armstrong, 2014).

These processing dynamics interact with the representations of ambiguous and unambiguous words to generate a range of different effects, which are described here relative to a baseline consisting of unambiguous words. The processing of unambiguous words consists of activating a single set of features that correspond to a distributed representation of the word’s semantic representation (Hinton & Shallice, 1991). Given that these features form a single coherent interpretation, all of these features would cooperate to activate one another early on; this will be referred to as a “moderate” amount of cooperation. Given the consistency of the interactions within semantics relative to the arbitrary nature of the mappings between the surface forms of words and their meanings, the contribution of the intra-semantic dynamics to the overall settling dynamics is assumed to be much stronger than the orthographic-semantic dynamics (for additional discussion, see Plaut & Shallice, 1993). Later, competitive dynamics are less relevant to unambiguous words as their bottom-up processing evokes no inconsistent semantic information.

Polysemes are assumed to be associated with two or more distinct but related senses that have considerable overlap in their semantic features. This assumption is consistent with the bulk of current ambiguity research (e.g., Klepousniotou et al., 2008; Rodd et al., 2002; Williams, 1992; but for a range of alternatives, see Klein & Murphy, 2001; Lehrer, 1990; Nunberg, 1979). As a result of these partially overlapping representations, polysems benefit from a high degree of cooperation across their related senses early on in processing, when processing is dominated by excitatory dynamics. This is because a larger set of features can cooperate to activate the core semantic features that are shared across
interpretations compared to an unambiguous word, thus giving rise to an early polysemy advantage, primarily driven by strong activation in the core features (a dynamic that is also consistent with the contextual underspecification hypothesis, Frisson, 2009). Later on, competition between the inconsistent aspects of the word’s senses will slow processing and eliminate this advantage. In the absence of systematically biasing context, the competition will be resolved on the basis of the relative frequency and richness of the alternative senses, perhaps also influenced by pre-existing semantic activation from prior (unrelated) lexical processing.

Homonyms are assumed to be associated with multiple unrelated—and therefore non-overlapping—semantic representations. Consequently, homonyms do not receive any cooperative advantage for processing early on, beyond that present within each of their individual interpretations. Homonyms therefore elicit similar initial semantic activation as unambiguous controls, albeit perhaps distributed more widely (e.g., twice as many features activated half as strongly via bottom-up input from the word form). As features associated with the incompatible meanings become more active, and as slower/weaker inhibitory dynamics increasingly influence later processing, competition among conflicting meanings increases. This competition is greater than it is for polysemous words, which have fewer inconsistent features, slowing processing and giving rise to a late homonym disadvantage (given that, late in processing, polysemous and unambiguous words are comparable). As with polysemes, the competition is ultimately resolved by the relative strengths of the alternative meanings and—when these are matched—random influences from prior activation.

4.2 Effects of biasing context on processing ambiguous words

As made clear in the review of the literature, early processing is largely insensitive to context. Thus, the main effects of context are on later, competitive dynamics. For unambiguous words, context may be slightly facilitatory because the word will generally be consistent with the context in which it occurs. However, this facilitation will not be very large because sufficient information to activate the semantic representation is available from the word form itself.

The effects of context are more substantial in the case of ambiguous words. We assume that context is sufficient to bias the competition within semantics so that the contextually appropriate sense or meaning eventually becomes fully activated and that, in a neutral context, the more dominant interpretation is derived. Thus, biasing context can either speed or slow processing depending on the relative dominance of the contextually appropriate interpretation. Compared to a neutral context, a context favoring the dominant interpretation will facilitate processing because the context provides additional support for the same interpretation. On the other hand, a context favoring the non-dominant interpretation will slow processing because, now, context must overcome competition from the contextually-inappropriate dominant interpretation. These effects will generally be greater for homonyms than for polysemes. Note that, in the latter case, the competition is only partial because some of the features of the inappropriate interpretation will be shared with the fully-activated appropriate interpretation.

The time point at which a substantial context effect will be observed will depend on the degree to which context biases the activation of one interpretation of an ambiguous word over another. Extremely biasing context (e.g., Tabossi, 1988) may, in fact, lead to a virtually immediate influence of context from the onset of semantic processing, although such strong biases are likely quite exceptional. The presentation of contexts that are unrelated to an ambiguous word may also lead the partially-activated semantic representation to move toward the interpretation that happens to be most supported by the unrelated context, even if this support is weak (Rodd et al., 2004).

In principle, the settling dynamics account should produce a range of effects at different points in time that are consistent with a broad set of reported findings, all based on domain-general connectionist principles. Indeed, it is worth briefly noting that, in several respects, the proposed processing dynamics share many similarities with connectionist accounts of other phenomena, such as the basic-level category label advantage (e.g., basic-label “dog” vs. superordinate-label “animal” or specific-label “collie”) and the facilitative or inhibitory role of context, as might be expected from an account couched in a domain-general computational formalism (Chen & Mirman, 2012, 2015; McClelland & Rogers, 2003; Rogers & Patterson, 2007; Tyler, Moss, Durrant-Heatfield, & Levy, 2000). For instance, in relation to the basic-label advantage, the polysemes share their core properties across all of their interpretations, which should lead to an initial advantage for learning the core features of a polyseme as well as a processing advantage as all of a polyseme’s features cooperate to activate the core features (Rogers & Patterson, 2007; see Tyler et al., 2000, for a model of category specificity that operates according to similar principles). Later on, competition between the inconsistent aspects of the word’s senses will slow processing and eliminate this advantage, similar to why specific labels are processed more slowly to basic-level labels. Similarly, in models of neighborhood effects in lexical processing (Chen & Mirman, 2012, 2015), early on the presence of neighboring representations generates a facilitative effect for the features shared across all neighbors, whereas later on the inconsistencies among neighbors generate a net inhibitory effect (see also Kawamoto, 1993; Kawamoto et al., 1994; Piercey & Joordens, 2000). Consequently, there is good reason to believe that the computational principles underpinning our proposed account have
the potential to produce the appropriate patterns of ambiguity effects at different points in time.

Even so, it is important to go beyond this verbal account to show that the relevant principles actually give rise to the empirically observed patterns of results. Accordingly, we carried out a small-scale simulation to examine how the cooperative and competitive dynamics within a biologically-motivated connectionist network interact with representational differences between polysemous, homonymous, and unambiguous words to produce different ambiguity effects at different points in time (see Armstrong & Plaut, 2008, for related work using a standard connectionist network).

### 3.3 Simulation

In order to elucidate the underlying mechanisms in the clearest possible manner, various aspects of the model were simplified as much as possible. Thus, the work should not be interpreted as a complete implementation of all aspects of the SSD account, but as a basic validation of its core principles. Full methodological details are given in Appendix , along with additional analyses of the learned connectivity structure that add further support for the principles of the account; here we present only the most theoretically relevant aspects of the simulation.

The simulation is instantiated in a more neurobiologically realistic variant of the standard connectionist formalism (see also Armstrong & Plaut, 2013; Laszlo & Plaut, 2012; Laszlo & Armstrong, 2014; Plaut & Booth, 2000; Watson, 2009) with the following properties: 1) there are distinct populations of excitatory and inhibitory units, as reflected by constraints on the sign of the unit’s outgoing weights; 2) there are far fewer inhibitory units than excitatory units; 3) the connections between layers (e.g., orthography and semantics) are only excitatory, whereas connections within a layer are both excitatory and inhibitory; and 4) the density/strength of between-layer connections is weaker compared with within-layer connections. Each of these properties contribute to encouraging stronger/faster excitatory dynamics and weaker/slower inhibitory dynamics.

The model architecture is presented in Figure 2. Four input units represent distinct orthographic word forms, whereas two others represent separate contexts in which these forms occur. These units are fully connected to semantic units with excitatory-only connections. The semantic units send excitatory-only connections to each other (excluding self-connections) and to an inhibitory unit, which projects back to semantics with inhibitory-only connections. Between-layer connections were subject to weight decay to limit bottom-up influences relative to within-layer dynamics.

The training set consisted of two unambiguous words, one homonymous word, and one polysemous word, each presented in two contexts. The meaning of each unambiguous word consisted of a fixed set of four distinct semantic features, regardless of context. For the homonym, two distinct sets of four semantic features were associated with the orthographic input, depending on which context was activated. For the polyseme, three semantic units were associated with the orthographic input regardless of the context but the fourth semantic unit was different for each context. The relatively weak/late effect of context was instantiated simply by activating the context unit a bit later than the orthographic input. The network was trained with a form of back-propagation (Pearlmutter, 1989) to activate the correct semantic features once both the orthographic input and context had been presented.

Figure 3 plots the sum of semantic activations that exceeded a threshold value of 0.5 after training. Examining the activation trajectories at successive time intervals, the model exhibits an early polysemy advantage (time A), followed later by a homonymy advantage and a polysemy disadvantage (time B), a homonymy disadvantage (time C), and finally, by a disadvantage for all ambiguous words (time D). An analysis of the pattern of learned weights in the network provides some insights into the cooperative and competitive dynamics that give rise to these patterns (see Appendix ).

Collectively, these results substantiate the verbal description of the SSD account in explicit mechanistic terms, in that the observed changes in semantic activation over time exhibit the core properties of the account. Experiments associated with fast responses, such as visual lexical decision (e.g., Beretta et al., 2005; Klepousniotou & Baum, 2007; Rodd et al., 2002), could reflect the dynamics near time A in Figure 3, leading to a polysemy advantage; extremely easy versions of lexical decision with very fast responses would tap semantics before this point when there is no substantial activity for any word class (Azuma & Van Orden, 1997; Rodd et al., 2002). Tasks that are associated with slower responses, such as semantic categorization involving broad categories (Hino...
3.4 Testing predictions of the account

The SSD account, as instantiated in the model, predicts that varying processing time alone, even if holding the task constant, should generate different ambiguity effects. In particular, conditions with a difficulty between standard (visual) lexical decision and semantic categorization with broad categories should produce both a weak homonymy disadvantage and a weak polysemy advantage (Figure 3, time B), in conjunction with mean overall latencies that fall in between those observed in those tasks. More generally, holding task (and the basic configuration of the decision system) constant, while varying overall latency of response is predicted to modulate semantic ambiguity effects in a specific, empirically testable manner. Here we review some preliminary efforts to test these predictions.

3.4.1 Effects of task difficulty in visual lexical decision.

In an initial study, we (Armstrong & Plaut, 2011) varied task difficulty in visual lexical decision with the aim of shifting from the standard polysemy advantage under easier (faster) conditions to both a polysemy advantage and homonymy disadvantage under more difficult (slower) conditions. Difficulty was manipulated by a combination of nonword difficulty (orthographically hard nonwords vs. orthographically very hard nonwords vs. pseudohomophones) and stimulus contrast (“full” white-on-black text vs. “degraded” gray-on-black text). Word stimuli were selected to fill a 2 (meanings: one vs. many) x 2 (senses: few vs. many) factorial design analogous to that used by Rodd et al. (2002), and which contain unambiguous, homonymous, polysemous, and hybrid ambiguous items, respectively.

We found that stimulus contrast slowed latencies much more than did nonword difficult, and thus we expected a only a polysemy advantage in the full contrast conditions, but a homonymy disadvantage along with a weak or absent polysemy advantage in the degraded conditions. A series of analyses revealed effects that were generally (although not perfectly) consistent with the predictions of the SSD account, provided that the statistical assessments took into consideration the relative meaning frequency of the homonyms, as predicted by other work (Armstrong, Tokowicz, & Plaut, 2012; Twilley, Dixon, Taylor, & Clark, 1994). There was, however, some inconsistency with which the homonymic and polysemic effects were modulated as a function of task difficulty, perhaps due to the between-participants nature of the task and the use of too many unbalanced homonyms. There may also be a limit on the degree to which visual lexical decision can be made to rely on higher-level (semantic) representations.

3.4.2 Comparing auditory versus visual lexical decision.

Rodd et al. (2002) found that, unlike visual lexical decision, auditory lexical decision produced a strong homonymy disadvantage and approximately 300 ms slower latencies (see also Mirman et al., 2010). Although these results are consistent with the predictions of the SSD account, the two tasks differed in the exact stimuli used—only about two thirds of the visually presented items appeared in the auditory version of the task, and so a direct comparison of item latencies in both tasks is not possible. Additionally, because the experimental work was conducted in English, it is difficult to ensure that other psycholinguistic properties, such as orthographic and phonological neighborhood sizes, are well matched across the different conditions.

To examine this issue more directly, we conducted additional analyses of the auditory and visual lexical decision data collected by Armstrong et al. (2014). The study was conducted in Spanish, which has much greater spelling-sound transparency than English and, hence, reduces the likelihood of confounds in orthographic versus phonological neighborhood sizes. This experimental work was originally designed to determine whether stronger and more complete
semantic settling occurs in the auditory modality relative to the visual modality at the time at which a lexical decision is made. The manipulation of modality was also conducted within participants in a blocked design, which should increase the power of the statistical tests between the two conditions. The initial results of that work suggested that, for a standard measure of semantic richness—imageability—as well as for a new measure based on the “ease of verbalizing” a word’s meaning, additional processing had indeed taken place in the auditory modality.

Here, we extend these results to the study of semantic ambiguity by analyzing effects of number of meanings and number of senses as covariates in the original experimental dataset (for additional details, see Armstrong et al., 2014; Barreiro Abad, 2014). Although the experiment was not initially designed to study semantic ambiguity per se, and Spanish has even fewer balanced homonyms than English (Armstrong et al., in press), the data set nevertheless contained 14 homonyms, 24 polysemes, and 350 unambiguous words.\(^7\) The number of ambiguous words is thus still comparable to other published work (Hino et al., 2010; Mirman et al., 2010). Using mixed-effect regression (Baayen, Davidson, & Bates, 2008) and following a similar analytical approach to Armstrong and Plaut (2011), the only comparison relative to unambiguous words that was significant was a homonymy disadvantage in the auditory condition (all \(p\)’s ≥ .28). This finding is consistent with the predictions of the SSD account and with prior related studies of auditory lexical decision (e.g., Mirman et al., 2010; Rodd et al., 2002), but in a more controlled language for clearly comparing auditory versus visual conditions. Additional empirical work is, however, needed using an auditory versus visual lexical decision paradigm that is explicitly designed to contrast ambiguity effects to fully establish this effect.

In summary, these initial empirical studies, although clearly in need of follow-up work to address various aspects of the experimental designs, offers some initial promise that the unique predictions of the SSD account do manifest in the empirical data. Additional work using other paradigms (e.g., response deadline tasks; Rogers & Patterson, 2007) will be necessary, however, to draw stronger conclusions in this respect.

4. General Discussion

A full review of the literature on ambiguous word processing, both in isolation and in biasing context, reveals complex and often apparently contradictory effects of number and relatedness of interpretations. Broadly speaking, in tasks with shorter latencies (e.g., visual lexical decision), one tends to observe a polysemy advantage but little if any homonymy disadvantage (each relative to unambiguous words). In tasks with longer latencies and/or which emphasize semantic processing (e.g., auditory lexical decision, semantic categorization) the pattern reverses, with a weaker polysemy advantage but now a strong homonymy disadvantage. Context effects are generally weak earlier in processing, allowing multiple interpretations to be co-active, with the timecourse of deactivation of the inappropriate interpretation varying as a function of contextual constraint and interpretation dominance.

These findings poses a major challenge for theories of word comprehension, and discourse comprehension more generally. To date, most theories have been narrow in the scope of issues that they address and/or not specified in explicit computational terms. This has cast doubts on the explanatory value of those theories and has limited their ability to provide targeted predictions that guide empirical work (McClelland, 2009).

To improve this state of affairs, we developed the semantic settling dynamics (SSD) account of the time-course of ambiguous word processing. This account posits that neurobiologically plausible excitatory and inhibitory processing dynamics interact with the representations of the number and relatedness of a word’s interpretations to cause different types of ambiguity effects at different points during processing. The SSD account explains a broad scope of existing empirical findings, and generates targeted predictions for guiding coordinated empirical research. Promising initial computational and empirical work also lends some preliminary support for the unique predictions of the theory that are not shared by other competing accounts. The account is, of course, incomplete, and this review highlights several directions for valuable follow-up investigations.

More sophisticated computational simulations could serve to better develop some of the simplified aspects of the model and test their ramifications via coordinated empirical investigations. For instance, giving the model the computational capacity to process a stream of text and learn to extract its own internal contextual representations will allow for more detailed study of how context that precedes or follows an ambiguous word shapes how comprehension unfolds over time (see Frazier & Rayner, 1990; McClelland et al., 1989; Rodd et al., 2013). Larger vocabulary sizes, elaborated semantic representations, and ranges of relative meaning frequencies (see Griffiths, Steyvers, & Tenenbaum, 2007; Hoffman et al., 2013; Johns & Jones, 2010; Plaut, 1995) will also enhance the model’s ability to make tight connections with many empirical phenomena, such as the details of how competition effects are modulated by relative meaning frequency and how differing degrees of relatedness of interpretation influence polysemy effects. In addition, integrating lexical-level modeling with more detailed instantiations of frontally-mediated systems such as working memory and ambiguity resolution.

\(^7\)An additional 10 ambiguous words were dropped because they had an atypically large number of senses—22 on average—relative to the rest of the data set and to most past work (e.g., Armstrong & Plaut, 2011; Klepousniotou & Baum, 2007; Rodd et al., 2002).
Incorporating additional neurobiological constraints that allow for the simulation of electrophysiological correlates of semantic processing would allow for a more transparent mapping of the time-course of processing in the model to the time-course of processing in humans, thus connecting the SSD account even more directly with that literature (e.g., Laszlo & Plaut, 2012; Laszlo & Armstrong, 2014). This is particularly important because of the apparent paradox that the presence of prior, reasonably strongly constraining context presents for broader theories of context integration. For instance, many researchers have offered theoretical accounts of discourse processing that involve the generation of strong forward predictions regarding the upcoming content of the input stream (e.g., Altmann & Mirković, 2009; Kukona, Fang, Aicher, Chen, & Magnuson, 2011). Those accounts would predict that when disambiguating context precedes an ambiguous word, there would be no reason for the incorrect interpretation of an ambiguous word to be activated at all. Nevertheless, the effects of context on semantically ambiguous words, with the possible exception of extremely constraining contexts, are inconsistent with this prediction. Possible explanations for this paradox which would benefit from explicit simulations include that context may not generally be quite as predictive as has previously been theorized, or that the initial bottom-up wave of activation from the stimulus itself is still sufficient to transiently activate the inconsistent interpretation. This latter effect may, indeed, be an emergent property of neural networks trained to recognize words in natural discourse. For example, occasional garden path sentences (e.g., “John was out of money and needed to go to the BANK to catch his dinner”) may bias the activation of what are ultimately contextually incorrect interpretations, as may the need to extend new interpretations to known words (Frison & Pickering, 2007). The presence of these types of stimuli in natural language may therefore encourage the word comprehension system to allow for some degree of temporary activation for an interpretation that is likely contextually inconsistent, so as to be able to resolve these types of sentences.

Developing this account within the context of a framework that incorporates learning mechanisms as a core feature also opens up interesting avenues for relating the SSD account not only to typical performance in college age students, but throughout the development of linguistic abilities, as well. For instance, several empirical investigations have now studied the time-course with which the interpretations of ambiguous words are activated, as well as how long those interpretations remain active, at different points throughout development (Booth et al., 2006; Khanna & Boland, 2010; Simpson & Foster, 1986; Simpson, Lorsbach, & Whitehouse, 1983). These studies indicate that, early on in development, children (age 5-7 years) initially activate all of the meanings of known homonyms regardless of the context in which the words are encountered (Khanna & Boland, 2010), and both meanings remain active for an extended period of time (Booth et al., 2006; see also Cairns et al., 2004, for a discussion of whether children maintain the activation of both meanings from the onset of reading acquisition, or if they initially tend to commit to one meaning and later learn to activate and maintain both). Inhibitory processes that suppress the contextually inappropriate meaning develop more slowly and do not reach mature levels until much later (9-12 years). Consequently, contextually inappropriate meanings that are usually suppressed by adults in the first 500 ms of processing may take closer to 1000 ms for children to suppress, if they are suppressed at all. These results would appear to provide important constraint and predictions for how the inhibitory mechanism in the SSD account should develop if the model was tested at different points during learning. Given that the developmental literature, has, more broadly, shown that a child’s ability to detect and think about ambiguous word meanings, in a metalinguistic sense, are both predictors of overall reading comprehension (Cairns et al., 2004; Zipke, Ehri, & Cairns, 2009), targeted insight into the specific mechanisms of ambiguous word comprehension may reveal important new avenues for understanding language development and, ultimately, improving language instruction and the remediation of language impairments. Given the similarity between the timecourse of ambiguity effects early in development and those observed in a range of different neural impairments often associated with brain damage later in life, as discussed in the initial review, the insights from early development may also have important implications for neural impairments, particularly to inhibitory processes, as well.

The study of the inhibitory mechanics incorporated into the SSD account may also have wider-spread implications for theories of lexical processing and for domain-general theories of inhibitory control (e.g., de Bruin, Roelofs, Dijkstra, & FitzPatrick, 2014; Green, 1998) which warrant further study in their own right. In particular, the bulk of the ambiguity effects reported in the literature have indicated that the contextually inappropriate interpretation of an ambiguous word is inhibited to the same extend—but not more than—control items. These results, however, are inconsistent with theories of negative priming in which inconsistent items, broadly construed, are inhibited below baseline (Gernsbacher & Faust, 1991; Houghton & Tipper, 1996; Tipper & Driver, 1988). Understanding the cause of these discrepancies may therefore shed important light on the general-versus-specific nature of inhibitory processes and the types of task demands and representations that can lead to these different patterns of results. Interestingly, the timecourse of inhibition may be relevant to reconciling these differences, just
as the temporal dynamics associated with semantic processing were for explaining many apparently inconsistent results in the semantic ambiguity literature. This possibility is supported by results reported by Nievas and Mari-Beffa (2002), who aimed to determine whether task demands and degree of contextual bias could have underpowered the detection of the effects of selective inhibition. In a primed lexical decision task, they observed that strong primes were, in fact, able to generate small but statistically significant amounts of selective inhibition. Moreover, detailed analyses of participants’ speed-accuracy trade-offs showed that this selective inhibition effect increased as more time passed—average-to-fast responders did not show a substantial inhibitory effect, but the slowest responders with the highest accuracy did. 8

Nievas and Mari-Beffa interpreted these results as evidence for selective inhibition, likely resulting from inhibitory control. A similar conclusion was also drawn by Reder (1983), who reported negative priming for the contextually-inconsistent interpretation of a word, as assessed by interpretation errors. Clearly then, further investigation of this issue promises to advance not only the SSD account, but also our understanding of inhibition, as well. Nevertheless, it is important to keep these effects in context given the general failure to find effects of selective inhibition in the literature and the fact that when such effects are detected they are typically weak, occur late in time, and require extremely constraining contexts. Thus, the more critical take-home message from these data for the purpose of a theory of ambiguity effects based on the timecourse of processing is that inhibitory processes are generally slow, and/or weak, an relatively non-selective, as described in the current formulation of the SSD account.

On another front, combining our account of ambiguity effects in the semantic system with a detailed specification of the representations and processes in orthography and phonology (e.g., McClelland & Elman, 1986; Plaut, McClelland, Seidenberg, & Patterson, 1996; Rumelhart & McClelland, 1986) could provide additional insights into lexical processing. For example, staged models of written word comprehension (e.g., Borowsky & Besner, 1993, 2006) predict that stimulus degradation would extend visual/orthographic processing but not alter the amount or precision of semantic processing. Armstrong and Plaut’s (2011) report of modulation of ambiguity effects as a function of stimulus contrast challenge such an model and support a cascaded theory of lexical processing, more in keeping with standard processing assumptions in neural/connectionist networks. Moreover, the re-analysis of the auditory versus visual data reported by (Armstrong et al., 2014), as well as the prior related reports by Rodd et al. (2002) and Mirman et al. (2010), suggest that auditory comprehension may involved faster cascaded processing than visual comprehension. Thus, the way in which information flows from orthographic or phonological representations to semantics may be shaped in non-trivial ways by other factors, such as whether information enters the system simultaneously versus sequentially in the visual versus auditory word recognition systems, respectively. Additional computational work is needed to better understand whether these more detailed semantic settling dynamic effects are captured by classic models that have focused on processing a perceptual input up to a “lexical” level of representation, if they are expanded to include semantic representations, as well (e.g., McClelland & Elman, 1986; McClelland & Rumelhart, 1981). In particular, it will be interesting to determine whether additional, more biologically-plausible computational principles are needed to explain these effects (for related discussion, see Carreiras et al., 2013). For practical purposes, the present results, when considered in conjunction with previous findings (e.g., Chen & Mirman, 2012; Mirman et al., 2010; Rodd et al., 2002) also suggest that an exploration of the temporal dynamics underlying semantic processing in general may be well-served by additional studies in the auditory modality, provided that appropriate stimulus constraints are employed (for discussion, see Rodd et al., 2002).

More broadly, the present results also call for a tighter integration of theories of lexical processing with theories of response selection. Although the SSD account was developed originally as a direct alternative to an account based on qualitative differences in how decisions are made across tasks—and without retracting the present claim that semantic settling dynamics explain most of the relevant findings—it is clearly an oversimplification to ignore the contributions of the response system in determining how and why a response is made. The development of such an integrative account will not only allow for more direct contrasts of where the relevant dynamics that underlie a given empirical phenomenon are occurring, but also for a more compelling understanding of these systems as an interactive network. For instance, it remains to be formalized in explicit computational terms how it is determined when sufficient information—semantic or otherwise—has been accumulated to reliably make a response in a given task—or indeed, within different conditions of a single task (see Armstrong & Plaut, 2011).

8Jastrzembski and Wittes (1982) reported a strikingly different pattern of results, such that an overall ambiguity advantage was detected for fast responders but not for slow responders. However, no distinction between homonyms and polysemes was made when selecting their ambiguous words. Based on the extremely high number of meanings (33) reported for their “high-number-of-meanings” items and the relative paucity of homonyms with more than 2 meanings in English (Armstrong, Tokowicz, & Plaut, 2012), it seems likely that most of their ambiguous items were, in fact polysemes; in which case, their results are consistent with the observation of stronger polysemic effects for tasks with faster responses and weaker or absent polysemic effects for tasks associated with slower responses.
Along these lines, whereas word latencies reported by Armstrong and Plaut (2011) typically were not drastically altered by increases in nonword wordlikeness, the nonword latencies increased to a much larger extent. The pseudohomophones were also associated with the largest increase in both word and nonword latencies but did not produce semantic effects consistent with more resolved semantic representations. Both of these effects are difficult to reconcile with an account based strictly on semantic settling dynamics but would fall out naturally from an extension of that account that incorporated a response selection system as well. For instance, the slowing of only one type of response suggests that other aspects of the cognitive system, including the decision system, are adapting to the change in stimuli, thereby reducing the predicted effects on the word condition. Moreover, the decision system may learn to tap different types of representations to gather the evidence needed to make lexical decisions in the context of different types of nonwords (e.g., pseudohomophones). Indeed, perhaps because pseudohomophones are able to engage specific semantic representations via phonology, the response system may learn to de-emphasize evidence from semantics because it is less informative in separating words from nonwords. The longer latencies in that condition may, therefore, be due to the response system needing to wait for additional non-semantic information, such as precise orthographic information, to be resolved (Hino & Lupker, 1996). A similar principle may underlie the results reported by Hino et al. (2010), in which Kanji nonwords composed of characters with meanings similar to the words weakened some ambiguity effects despite increasing overall latencies. Collectively, these findings emphasize the need for more integrative theories and explicit models that flesh out the specific contributions of the lexical system and the response system (for discussion, see Armstrong, Joordens, & Plaut, 2009; Armstrong & Plaut, 2013).

Presenting a more comprehensive model of specific tasks that includes both a lexical system and a response system is not a trivial undertaking, however, given the contention surrounding how the response system should be integrated with the lower-level systems that provide evidence to it (for discussion, see Armstrong & Plaut, 2013; Ratcliff, Van Zandt, & McKoon, 1999; Usher & McClelland, 2001). The empirical data—and in particular, performance in the pseudohomophone condition—also suggest that semantics, the response system, and other representations such as orthography and phonology interact in highly complex ways. More sophisticated modeling involving a full orthographic-phonological-semantic lexical processing model, more realistic context and discourse processing considerations, and an explicit response system will clearly be needed to fully understand these data.

5. Conclusion

The semantics settling dynamics account, although an im-
portant step in its own right, is still clearly in its infancy. By being couched in domain-general learning and more neurobiologically-plausible processing principles, it has considerable promise to provide a comprehensive account of semantic ambiguity effects at the lexical level, to connect to a broad array of types of behavioral and neural data, as well as to a broad set of phenomena in other related domains, such as discourse comprehension and response generation. The SSD account should thus continue to make valuable contributions to our understanding of these issues and beyond in future work.

References


Appendix: Simulation Details

A.1 Network architecture. The model architecture is presented in Figure 2. In keeping with its minimalist design, the network was composed of four orthographic input units, each of which was used to represent an individual word, two context units, each representing one of two separate contexts, and 22 units representing semantics. Of these 22 units, 21 were excitatory units and were used to code the distributed semantic representations of the words presented during training. The remaining unit was inhibitory and did not correspond to an explicitly-specified semantic feature.

Both the orthographic and the context units were connected to the semantic units. These between-layer connections were restricted to be excitatory only (non-negative). The excitatory semantic units were connected to one another with excitatory connections, with the exception that units did not connect to themselves. All of the excitatory weights were initialized to a value of 0.05, such that the initial magnitudes of the orthographic and context inputs were equal. Thus, any differences that emerge between the orthographic and context weights should be due to how critical each type of information is to activating the representation of each word class. The inhibitory weights were initialized to a value of -0.2. This balanced, in approximate terms, the amount of excitation that a semantic unit was expected to receive early in training with an equivalent amount of inhibition. Additionally, all of the units in the network had a bias connection—equivalent to a connection from an additional input unit that is always active—that was set to an initial value of 0. There was no other variability in the weight values to simplify the interpretation of the weight structure of the network. To constrain the magnitudes of the weights that could be learned for connections between versus within layers, the between-layer connections were subjected to weight decay after each weight adjustment (λ = 0.00025). Within-layer connections, including the connections to and from the inhibitory unit, were not subject to any weight decay.

A.2 Training patterns. The training representations used in the simulation are presented in Table 1. Localist representations were used to encode each word form and context. A distributed semantic representation was associated with each word/context input pair, and different groups of units were used to represent the semantic features associated with each interpretation of a word.

The training set consisted of two unambiguous words, one homonymous word, and one polysemous word. Each unambiguous word consisted of the pairing of a single orthographic unit with a group of four active semantic units. These
orthographic/semantic input/output pairs were presented in two different contexts, represented as the activation of one of the two context units. The pattern structure for ambiguous words differed as follows: For the homonyms, two different groups of four semantic units were associated with the orthographic input, depending on which context was activated. For the polysemes, three semantic units were associated with the orthographic input regardless of the context but the fourth semantic unit was different for each context.

The presentation of each word was structured so that the orthographic input would be activated for 20 unit updates, after which both the orthographic and context outputs were activated for an additional 20 updates. This was intended to reflect the weak/late effect of context in as simple a way as possible and is not to be interpreted as implying a formally staged theory of contextual access (cf. Sternberg, 1969; Swinney, 1979).

**A.3 Training.** The network was trained using recurrent back-propagation through time (Pearlmutter, 1989) and momentum descent. A learning rate of 0.0001 and momentum of 0.9 were employed (momentum was set to 0 for the first update). A relatively small learning rate was selected to maintain a positive gradient linearity across the weights that resulted from each weight update, which is more difficult if excitation and inhibition are represented in distinct pools of units. Units were considered to be correctly activated when they were within 0.15 of their target activation. Error was computed for the last 5 unit updates. Between each training pattern, the net input was reset to -1.4 and the net output was reset to 0.2 for all of the units. Weights were updated after each sweep through the training vocabulary. The model was trained until all semantic units were within 0.5 of their targets for each of the input patterns.

**A.4 Weight structure.** The weight structure of the network was examined to determine how the network’s knowledge for these words, as encoded in the weights, could interact with the representations of the different words to generate the observed semantic settling dynamics. Of course, the weights only form a part—albeit a very important one—of understanding the full complexity of the settling dynamics within the network. Note that because there was no random variance in the initial weights, the reported values are identical for all units within each sub-network.

Within semantics, the analysis revealed strong excitatory sub-networks for each interpretation of a word. Numerically, the weights among the core features were higher for the polyseme (weight value = 2.93) than for either the homonym or the unambiguous words, which had approximately equal intra-semantic excitatory weights (2.38 and 2.37, respectively). The distinct features of the polyseme had weaker incoming weights from the core features of the polyseme (1.08), but sent relatively stronger outgoing weights to the core features (1.23), and were disconnected from the distinct feature with which they were not consistent (0.0); the same was true for the connectivity between the semantic features used to code each meaning of the homonym. The learned weight structure associated with the polysemes is also similar to that which underlies basic-level category organization and the associated basic-level processing advantage (Rogers & Patterson, 2007).

The input from the inhibitory unit was largest for the homonym’s features (-2.21), followed by the input to the core features of the polyseme (-2.00) and unambiguous words (-1.87); the least inhibition was associated with the distinguishing features of the polyseme (-1.66). These results are consistent with the predictions of the account, in terms of how much competition should exist for homonyms (high competition), polysemes (low competition), and unambiguous words (no competition). The fact that the distinct features of the polyseme receive less inhibition is commensurate with the reduced excitation that they receive from the core features and the other distinct features.

The incoming weights from orthography and context were an order of magnitude smaller than the weights within semantics, consistent with the assumption (and imposed model architecture) that such connectivity better reflects the actual connectivity structure in cortex. Feeding into semantics, the orthographic inputs with the strongest weights projected to the core features of the polyseme (0.14), followed closely by the weights to the unambiguous words’ semantic features (0.13), by the weights to the semantic features of the homonym (0.10), and finally by the weights to the distinguishing features of the polyseme (0.07). With respect to the context units, the semantic features of the unambiguous words and the core features of the polyseme word were not sensitive to this information (weights = 0). The distinguishing features of the polyseme had moderate incoming weights from context (0.14), whereas the semantic units associated with the homonym received a relatively strong weight from context (0.25). These connectivity patterns are consistent with the differential need to rely on context to activate an appropriate semantic representation.

Overall, this connectivity structure is consistent with the predicted relative contributions of word form and context, and of cooperation and competition among consistent and in-
consistent semantic features, that underlie the SSD account. It also helps flesh out how the presentation of a given input representation can interact with the network’s memory, as encoded in the weights, to give rise to the simulated semantic settling dynamics.